

How to go to your page

This eBook contains multiple volumes. Each volume has its own page numbering scheme for the front matter.

For example, to go to page IV of Volume 1, type v1-iv in the "page #" box at the top of the screen and click "Go." To go to page IV of Volume 2, type v2-iv and so forth.

All other page numbers are contiguous between the volumes.

VISIT...

LANZAROTE
Caliente.COM

VOLUME 1

PHONOLOGIES
OF

ASIA
AND
AFRICA

Editor
Alan S. Kaye

Technical Advisor
Peter T. Daniels

Phonologies of Asia and Africa

This page intentionally left blank.

Phonologies of Asia and Africa

(Including the Caucasus)

Volume 1

edited by

Alan S. Kaye

Technical Advisor

Peter T. Daniels

Winona Lake, Indiana

EISENBRAUNS

1997

© Copyright 1997 Eisenbrauns

All rights reserved

Printed in the United States of America.

Library of Congress Cataloging in Publication Data

Phonologies of Asia and Africa : (including the Caucasus) / edited by
Alan S. Kaye ; technical advisor, Peter T. Daniels.

p. cm.

Includes bibliographical references.

ISBN 1-57506-017-5 (vol. 1: cloth : alk. paper). — ISBN 1-57506-018-3
(vol. 2: cloth : alk. paper). — ISBN 1-57506-019-1 (2-vol. set: cloth : alk.
paper)

1. Asia—Languages—Phonology. 2. Africa—Languages—
Phonology. I. Kaye, Alan S. II. Daniels, Peter T.

P381.A75P48 1997

414—dc21

97-4964

CIP

The paper used in this publication meets the minimum requirements of the
American National Standard for Information Sciences—Permanence of Paper
for Printed Library Materials, ANSI Z39.48-1984. ∞

Contents

Volume 1

List of Maps	ix
List of Tables	ix
Preface	xiii
Introduction	xv

1 Afroasiatic Languages

Semitic Languages

Ancient and Medieval

East Semitic

1	<i>Akkadian and Amorite Phonology</i>	3
	Giorgio Buccellati	
2	<i>Eblaite Phonology</i>	39
	Cyrus H. Gordon	

Central Semitic

3	<i>Ugaritic Phonology</i>	49
	Cyrus H. Gordon	

Northwest Semitic

4	<i>Phoenician and Punic Phonology</i>	55
	Stanislav Segert	
5	<i>Ancient Hebrew Phonology</i>	65
	Gary A. Rendsburg	
6	<i>Tiberian Hebrew Phonology</i>	85
	Geoffrey Khan	
7	<i>Jewish Palestinian Aramaic Phonology</i>	103
	Geoffrey Khan	
8	<i>Old Aramaic Phonology</i>	115
	Stanislav Segert	
9	<i>Classical Syriac Phonology</i>	127
	Peter T. Daniels	

vi Contents

- 10 *Modern and Classical Mandaic Phonology* 141
Joseph L. Malone

South Semitic

- 11 *Old South Arabian Phonology* 161
Gene Gragg
- 12 *Ge'ez Phonology* 169
Gene Gragg

Modern

Central Semitic

- 13 *Arabic Phonology* 187
Alan S. Kaye
- 14 *Moroccan Arabic Phonology* 205
Jeffrey Heath
- 15 *Cypriot Arabic Phonology* 219
Alexander Borg
- 16 *Maltese Phonology* 245
Alexander Borg

Northwest Semitic

- 17 *Israeli Hebrew Phonology* 287
Shmuel Bolozky
- 18 *Modern Aramaic Phonology* 313
Robert D. Hoberman

South Semitic

- 19 *La phonologie des langues sudarabiques modernes* 337
Antoine Lonnet et Marie-Claude Simeone-Senelle
- 20 *Chaha (Gurage) Phonology* 373
Wolf Leslau
- 21 *Amharic Phonology* 399
Wolf Leslau

Egyptian Sub-branch

- 22 *Egyptian and Coptic Phonology* 431
Antonio Loprieno

Berber Languages

- 23 *Berber Phonology* 461
Maarten G. Kossmann and Harry J. Stroomer

Cushitic Languages

- 24 *Awngi Phonology* 477
Robert Hetzron

25	<i>Oromo Phonology</i>	493
	Maria-Rosa Lloret	
26	<i>Somali Phonology</i>	521
	Annarita Puglielli	
Chadic Languages		
27	<i>Hausa Phonology</i>	537
	Paul Newman	

Volume 2

2 Indo-European Languages

Ancient and Medieval

Anatolian Languages

28	<i>Hittite Phonology</i>	555
	H. Craig Melchert	

Iranian Languages

29	<i>Old Persian and Avestan Phonology</i>	569
	David Testen	
30	<i>Pahlavi Phonology</i>	601
	Dieter Weber	

Modern

Indo-Aryan Languages

31	<i>Hindi-Urdu Phonology</i>	637
	Alan S. Kaye	
32	<i>Gujarati Phonology</i>	653
	P. J. Mistry	

Iranian Languages

33	<i>Persian Phonology</i>	675
	Gernot L. Windfuhr	
34	<i>Kurdish Phonology</i>	691
	Ernest N. McCarus	
35	<i>Ossetic Phonology</i>	707
	David Testen	
36	<i>Pashto Phonology</i>	733
	Josef Elfenbein	
37	<i>Balochi Phonology</i>	761
	Josef Elfenbein	

viii Contents

	Armenian Sub-branch	
38	<i>Armenian Phonology</i>	777
	John A. C. Greppin	
3	Dravidian Languages	
39	<i>Brahui Phonology</i>	797
	Josef Elfenbein	
4	Nilo-Saharan Languages	
40	<i>Nilo-Saharan Phonology</i>	815
	M. Lionel Bender	
5	Niger-Congo Languages	
41	<i>Swahili Phonology</i>	841
	Ellen Contini-Morava	
42	<i>Sango Phonology</i>	861
	James A. Walker and William J. Samarin	
6	Altaic Languages	
	Turkic Languages	
43	<i>Turkish Phonology</i>	883
	Bernard Comrie	
44	<i>Tatar (Volga Tatar, Kazan Tatar) Phonology</i>	899
	Bernard Comrie	
45	<i>Uyghur Phonology</i>	913
	Bernard Comrie	
7	Caucasian Languages	
46	<i>Georgian Phonology</i>	929
	Howard I. Aronson	
47	<i>Chechen Phonology</i>	941
	Johanna Nichols	
48	<i>Lak Phonology</i>	973
	Gregory D. S. Anderson	
8	Unaffiliated Languages (Language Isolates)	
49	<i>Sumerian Phonology</i>	1001
	John Hayes	
50	<i>Burushaski Phonology</i>	1021
	Gregory D. S. Anderson	

Preface

This work has evolved over the course of the past few years and has involved the cooperation of 39 distinguished linguists. As such, uniformity in methodology and terminology has not been a desideratum, nor would it have been possible to attain. The achievements, nevertheless, have, in my opinion, proved to be remarkable. The goal of this publication is to provide a summary of what is currently known about the phonology of selected Asian and African languages. In addition, three languages of the Caucasus are included. What I, as editor, have strived to do here is to give each author enough leeway to allow for the presentation of interesting facts in an informative way, while paying strict attention to what is most significant about the language(s) under investigation. Of course, the readership will ultimately have to take it upon itself to decide the extent to which I have succeeded. I would like to express my sincere thanks to all the contributors for their efforts, and above all, for their extraordinary patience while the book was in production. These collaborative projects often take much more time to complete than was originally envisaged, as I found out once before with the two volumes containing approximately 1,800 pages by 134 scholars I edited in 1991: *Semitic Studies in Honor of Wolf Leslau on the Occasion of His Eighty-Fifth Birthday*.

One contributor deserves special commendation. Peter T. Daniels has served admirably as the technical advisor to this collection. His service has proven to be exemplary and far beyond the call of duty. Let me briefly elaborate. Peter has been influential in the final product in three crucial ways. First, he performed many copy-editing tasks and was responsible, along with his Abjad Press, for the typesetting of the articles, often corresponding with authors on both technical and scientific matters. All the contributions are better off as a direct result of Peter's personal intervention. His commitment to excellence is obvious on each page which follows (and can also be seen in *The World's Writing Systems*, which he co-edited with William Bright and single-handedly typeset). Second, it was Peter who first suggested to me the idea that Tiberian Hebrew should be covered in a separate undertaking, in addition to the commissioned chapter on Biblical Hebrew. Readers could thus compare and contrast both these treatments. He further recommended

that since the Aramaic coverage was to be so extensive among all the Semitic languages discussed in the book, a special effort should be made to treat Palestinian Jewish Aramaic in an entire article by itself. As it turned out, both of his proposals came to fruition. Third, Peter recruited his colleagues at the University of Chicago, Howard Aronson, David Testen, and Gregory D. S. Anderson, to contribute to this reference by writing on their specializations, and also secured the services of Johanna Nichols to prepare the article on Chechen. These particular chapters have given this work, already a hefty tome prior to the arrival of this new material, an even greater appeal and strength by virtue of the variety of the additions, thus unquestionably justifying the larger size. Although I, as editor, bear full responsibility for the final selection of languages treated herein, it is Peter who must be credited for his insight that the volume should delve somewhat more into the heart of Africa than was originally contemplated, and offer as wide a coverage as possible, given our original framework of the prospective title, *Phonologies of Asia and Africa*.

Finally, I wish to express my gratitude to Wolf Leslau and Franz Müller-Gotama, who made a number of useful suggestions concerning the introduction.

Alan S. Kaye
Fullerton, California
December 1995

Introduction

Why yet another book on phonology when, apparently, there are already so many published volumes of every conceivable size, shape, orientation, theory, and methodology imaginable that deal with this subject? I hope to convince the reader that this book is unique, and furthermore, that it yields valuable insights to anyone interested in any type of descriptive phonology. In offering both introductory undergraduate as well as graduate courses in this field for almost 30 years, I have had the opportunity to use a variety of books and other materials. Most enjoyable for me and for the great majority of my students (as they have confided to me) have been articles (from a wide assortment of journals) that are devoted to specific languages. These include works by such notables as Edward Sapir, John Rupert Firth, Roman Jakobson, Zellig Sabbetai Harris, and numerous other outstanding scholars. These linguists have all had one thing in common—vast personal experience working on many languages and dialects, often via fieldwork and active language learning, be it in coursework or through independent study. All have introduced the reader to a large assortment of different perspectives—a virtual smorgasbord of aromas and tastes—everything from the Prague or London Schools (prosodic analysis) to the Copenhagen School of glossematics, structural or Bloomfieldian and Neo-Bloomfieldian to stratificational approaches, generative phonology of all persuasions including natural generative phonology, natural phonology, lexical phonology, autosegmental phonology, CV phonology, particle phonology, atomic phonology, dependency phonology, metrical phonology, experimental phonology, moraic phonology, and so on. While the dominant paradigms within phonology and phonological theory have been developing and, as seems only too obvious to report, rapidly changing over the past few years and even decades, the data-oriented discussions by the linguists of yesteryear seem to have held their own. Here we may cite some of the classic papers in the field, such as Leonard Bloomfield's "Menomini Morphophonemics" (1939), or Edward Sapir's "La réalité psychologique des phonèmes" (1933), available in English since 1949. Like good wine, they have all mellowed with age; like a good movie from the 1930s or 1940s, they are still enjoyable; and like a good book from

bygone days, one can still profit from them, while noting beauty, care in their preparation, and refinement, not to mention a superb elegance of phraseology and in the handling of minute details. In my view, they still have something worthwhile to say, and their contents are, even after all these years, germane. As many have claimed, not everything new or newer is necessarily better, and data-oriented linguistics (and one may add philology in general) can still contribute to the larger picture of the complicated workings of *parole*, *langue*, and even *langage*—to the grand synthesis or syntheses, if you will. There have been many other data-oriented articles which have appeared over the years. However, nothing similar to the compilation of articles I wanted to assemble and edit for the general linguist as well as the specialist has ever, to my knowledge, appeared until now.

The idea for this volume came about as I searched in vain for a book which would enable my students to gain a concrete familiarity of solid phonological work by subjecting them to the exposure of many of today's (hard-)working linguists who would concisely describe and comment on the phonological processes in and structures of languages which they have carefully scrutinized, both ancient or medieval and modern. In many cases, the authors chosen are the leading specialists on these languages (and often the language families to which they belong as well). As I have repeatedly emphasized in some writings and especially in my classes over the years at California State University, Fullerton, linguistics to me is, first and foremost, about languages—real, natural languages with all their messy details, including exceptions to this rule or that. Let me reiterate this very position here. This philosophy of science—or, perhaps better stated, this philosophy of how to do science—is a premise about how to do linguistics and about what linguistics is (and, invariably, is not), or, at least, this type or genre of linguistics. My *Weltanschauung* has been that linguists research languages and study structures—a simple enough supposition, or so it would appear (contra Noam Chomsky, who has repeatedly stated that “linguistics is *part of* psychology”; cf., e.g., his *Language and Responsibility*, p. 43). This seems even truer when one talks specifically about phonetics and phonology (rather than linguistics as a whole, which, I readily admit, is, *mutatis mutandis*, an interdisciplinary and multidisciplinary science). Phonology has, in fact, over the years continuously seen a steady stream of textbooks, handbooks, reference works, chrestomathies, and “surveys” of varying length and quality. Indeed our library shelves seem to be bulging to the breaking point with reprinted works and commentaries thereon, even commentaries on the commentaries, on

the various schools of phonology, phonologists, and phonological scope, analysis, explication, interfacing, etc. Moreover, numerous new books regularly appear promising what jacket blurbs usually push as new universal insights into phonological structure, phonological typology, phonological universals, or some such arena. As illustrative of the two aforementioned points, let me make mention of *Readings in Linguistics I and II*, edited by Eric P. Hamp, Martin Joos, Fred W. Householder, and Robert Austerlitz; while Bruce Hayes's *Metrical Stress Theory* is a current exploration of this popular phonological theory. Furthermore, Michael Kenstowicz's *Phonology in Generative Grammar* is a 700-page encyclopedic discussion of various theoretical concerns, and John Clark and Colin Yallop and John Goldsmith satisfactorily survey more or less the entire field (Clark and Yallop, *An Introduction to Phonetics and Phonology*; Goldsmith, ed., *The Handbook of Phonological Theory*). What the entire field of phonology has always lacked and is still lacking until now is a book with the present focus and scope, in which the subject matter is thoroughly packaged, consumed, and digested as a data-oriented, descriptive discipline, and does not merely serve as a hot-plate for a rehash of rules of various types and layouts with assorted labels and devices reflectant of different premises, styles, interests, or parameters.

The only volume known to me which, at least in part, attempted to cover the phonetics and phonologies of various languages is *Manual of Phonetics*, edited by Louise Kaiser, although there are brief phonological discussions in other works dealing with numerous languages, such as *The World's Major Languages*, edited by Bernard Comrie; George L. Campbell's two-volume *Compendium of the World's Languages*; or *International Encyclopedia of Linguistics*, edited by William Bright. Part D.b of the aforementioned Kaiser tome, however, dealt exclusively with languages, and one author therein (Wolf Leslau writing on the Semitic languages) has managed to contribute two articles in the same general area of specialization for this volume. The profound difference between the earlier study by this celebrated linguist and the present essays herein is one of detail(s), since the former was severely limited in space (pp. 325–29), and the author had little choice but to deal in generalities only.

The other coverages in the Kaiser *Manual* were devoted to Romance, Germanic, Slavonic, Finno-Ugrian, and African tone languages (this is the terminology employed in that work). The paper on the tonal languages of Africa was allotted a total of six pages (pp. 330–35, by A. Burssens). Unfortunately, that important section of the volume on specific languages was

removed in the book's second edition, which, according to the new editor, Bertil Malmberg, was a completely revised and expanded version of the prior *Manual* edited by Kaiser. Malmberg justified this decision to excise this material (after "a long discussion between the publisher" and himself, p. v) stating that the information devoted to "particular languages ... in a book of this kind seems to be of little interest, firstly because the choice of languages described is completely arbitrary, secondly because the space admitted for each language or language family does not permit any real presentation of its particularities" (ibid.). The present work is meant to alter, at least somewhat, this lamentable state of affairs which has existed far too long.

I have always wanted my students to come away from their courses in phonology with dual experiences. First and most important, they should have gained expertise with the phonological structure of real languages, both synchronically and, to a lesser extent, diachronically. Theory comes and goes, yet the facts of Arabic phonology, e.g., have not changed very much over the past half century or more. One still needs, even after all these years, to make reference to the problem of "emphasis" (pharyngealization or velarization) and the "emphatics," no matter what labels are *au courant*, what terms are trendy or in vogue, or what types of rules are fashionable or conventions in bloom. On the other hand, general phonology is hardly recognizable from the days in which Charles F. Hockett, Kenneth L. Pike, André Martinet, and Zellig S. Harris were among the household names of the 1940s or 1950s. Secondly, students should have become familiar with competing analyses while, at the same time, also being sensitive to different approaches to linguistic explications. This tome has been designed to allow readers, including seasoned veterans in the field, the opportunity to come to grips with the phonologies of many non-European languages. All the articles which follow, the collaborative efforts of several different phonological traditions, are written by experts who have considerable *savoir-faire* with languages. All of these linguists enjoy national and international reputations.

Let me now devote a few words to the difficulty encountered in the selection of languages. Reviewers will inevitably point out that this language should have been included or that that one was unnecessary, and thus could have been omitted. Since my own forte as a Semitist and an Afroasiaticist has been in the languages of parts of Asia and Africa, these geographical areas have, of necessity, been emphasized (see below for further details). It must be admitted that I at one time toyed with the idea of entitling this man-

ual *The Phonology of Oriental Languages*. That idea had eventually to be abandoned because the designation “Orient” as currently used is inexact, and colleagues, especially those in the humanities or social sciences, would not be receptive to the idea that it did not, in fact, deal with Chinese, Japanese, Korean, etc., certainly considered “Oriental” according to any current sense of the term. The present title is also somewhat ambiguous, I confess; yet no promise was ever intended by this choice that all languages of this region would be covered. Rather, the implication was and still is that we would be dealing with selected languages. A further more pragmatic factor had to do with the availability of certain authors who could meet our time frame.

Let me now turn to the specifics about which languages and dialects have been chosen for inclusion. The reader will surely have become aware of the extensive coverage given to the Semitic languages. Since some Arabic dialects (so-called) are, in my opinion, better designated separate and distinct Semitic languages in their own right, I made a special endeavor to recruit the leading authorities on Moroccan and Cypriot Arabic so that the contrasts would be noticeable when one compared these with the phonological facts discussed in the articles dealing with Maltese and Arabic (Modern Standard and Egyptian [Cairene]). The other modern Semitic languages treated herein include some of the most interesting (and most involved) of the diachronic developments from the Proto-Semitic stage. These are: Modern Israeli Hebrew, Modern (or Neo-) Aramaic, really a conglomeration of different languages (not dialects), six Modern South Arabian languages, Chaha (one of several distinct Gurage [Ethio-Semitic] languages), and Amharic, the national language of Ethiopia.

The ancient (including medieval) Semitic languages include: Akkadian (Assyro-Babylonian) and Amorite, Eblaite, Ugaritic, Phoenician and Punic, Biblical and Tiberian Hebrew, Palestinian Jewish Aramaic, Old and Biblical Aramaic, Syriac, Mandaic, Old (Epigraphic) South Arabian (also known as Epigraphic South Arabic), and Ge’ez (Classical Ethiopic).

The related Afroasiatic (Hamito-Semitic or Afrasian) languages include: Ancient Egyptian and Coptic, the Berber languages (not dialects), the Cushitic languages—Awngi, Oromo, and Somali—and the Chadic language, Hausa.

The culturally and geographically related (to the above) languages include: Sumerian (an isolate), Hittite, Avestan and Old Persian (Old Iranian), Pahlavi (Middle Persian), Persian, Ossetic, Kurdish, Pashto, Balochi,

Armenian (all Indo-European), Brahui (Dravidian), Turkish, Tatar, and Uyghur (Turkic sub-branch of Altaic).

Finally, there is a section of languages (the first two of which continue the aforementioned grouping) of interest to phonological typologists, which includes: Hindi-Urdu (Hindustani) and Gujarati (Indo-Aryan sub-branch of Indo-European), Burushaski (an isolate), Swahili (the Bantu sub-branch of Niger-Congo), Sang(h)o (a creole based on Ngbandi, which is Niger-Congo), Nilo-Saharan languages, and Georgian, Chechen, and Lak (South, North Central, and Northeast Caucasian respectively).

It should be noted that a serious attempt has been made from the outset to ensure that the chapters would be both understandable and useful to a linguist of any persuasion, while simultaneously anticipating that anthropologists, sociologists, historians, and other researchers could utilize the pages which follow for their own scholarly pursuits. For any shortcomings in this regard, it is I, as editor, who must assume full responsibility for the articles, including both of mine. In other words, the usual disclaimers apply.

Finally, there is one last item—a request of the readership. It would be greatly appreciated if comments could be sent directly to me suggesting improvements on any of the selections for the book's planned subsequent editions.

Alan S. Kaye
Linguistics Program
California State University, Fullerton
Fullerton, California 92634-9480
e-mail: akaye@fullerton.edu
Fax: 714-449-5954

References

- Bloomfield, Leonard. 1939. "Menomini Morphophonemics." *Travaux du cercle linguistique de Prague* 8: 105–15.
- Bright, William, ed. 1992. *International Encyclopedia of Linguistics*. New York: Oxford University Press.
- Campbell, George L. 1991. *Compendium of the World's Languages*. London: Routledge.
- Chomsky, Noam. 1979. *Language and Responsibility*. New York: Pantheon.

- Clark, George, and Colin Yallop. 1995. *An Introduction to Phonetics and Phonology*, 2nd ed. Oxford: Blackwell.
- Comrie, Bernard, ed. 1987. *The World's Major Languages*. London: Croom Helm; New York: Oxford University Press.
- Daniels, Peter T., and William Bright, eds. 1996. *The World's Writing Systems*. New York: Oxford University Press.
- Goldsmith, John, ed. 1995. *The Handbook of Phonological Theory*. Oxford: Blackwell.
- Hamp, Eric P., Martin Joos, Fred W. Householder, and Robert Austerlitz, eds. 1995. *Readings in Linguistics I and II* (abridged edition). Chicago: University of Chicago Press.
- Hayes, Bruce. 1995. *Metrical Stress Theory*. Chicago: University of Chicago Press.
- Kaiser, Louise, ed. 1957. *Manual of Phonetics*. Amsterdam: North Holland.
- Kaye, Alan S., ed. 1991. *Semitic Studies in Honor of Wolf Leslau on the Occasion of His Eighty-Fifth Birthday*. 2 vols. Wiesbaden: Harrassowitz.
- Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Oxford: Blackwell.
- Malmberg, Bertil, ed. 1968. *Manual of Phonetics*. Amsterdam: North Holland.
- Sapir, Edward. 1933. "La réalité psychologique des phonèmes." *Journal de psychologie normale et pathologique* 30: 247–65. English orig. in *Selected Writings of Edward Sapir in Language, Culture, and Personality*, ed. David G. Mandelbaum. Berkeley and Los Angeles: University of California Press, pp. 46–60.

This page intentionally left blank.

1 Afroasiatic Languages

This page intentionally left blank.

Chapter 1
Akkadian and Amorite Phonology
Giorgio Buccellati
University of California, Los Angeles

1.1. The graphemic base

1.1.1. *The writing medium*

Akkadian and Amorite are dead languages, in the specific sense that their speakers died out around 1600 B.C. (for Amorite) and 600 B.C. (for Akkadian). Our reconstruction of both languages is thus based exclusively on the written record, except for the inferences that may be drawn from the fact that they are related to other Semitic languages for which there are informants. The written medium, though rich in information, presents considerable limitations which must be taken into account. In addition, two other filters must be reckoned with, particularly in any discussion of phonology.

(A) The writing system was not originally developed for a Semitic language, but rather for Sumerian. The process of adaptation to Semitic was gradual and organic, and was not governed by a-priori linguistic considerations. Especially in the early periods, the scribes, who were conversant with Sumerian, maintained the basic graphemic oppositions which were best suited for Sumerian and subsumed under them a variety of Semitic oppositions. For instance, it has been suggested (Gelb 1961: 31–33) that Sumerian had an opposition between stops (without distinction of voice) and aspirates; this two-way graphemic opposition was used to render a three-way opposition, in Akkadian, between voiced, voiceless, and emphatics. (An interesting

AUTHOR'S NOTE: I wish to thank the editor of this volume, Alan S. Kaye, for his thoughtful comments on the first version of my manuscript, and for various substantive and bibliographical suggestions. ABBREVIATIONS: Am, Amorite; Ar, Arabic; OA, Old Assyrian; OAkk, Old Akkadian; OB, Old Babylonian; OB+, Old Babylonian and later dialects; PS or *, Proto-Semitic.—Standard symbols used in Assyriological literature: $\dot{1} = \dot{a}$; $\dot{2} = h$; $\dot{3} = \dot{h}$; $\dot{4} = \dot{s}$; $\dot{5} = \dot{g}$; $\dot{6} = w$; $\dot{7} = y$. Note also: $\underline{t} = \theta$. For $\dot{s}$, $\dot{s}$, and $\dot{g}$, see § 1.3.5. Square brackets are occasionally used in Assyriological literature to indicate the graphemic transliteration. This however generates confusion with the more common linguistic use of square brackets for phonetic transcription. Accordingly, where an explicit graphemic notation is needed, I either render transliterated cuneiform signs in small capitals (DA), or else I enclose them in angle brackets (<da>). Small capitals are also used to render logograms (generally, in their Sumerian reading).

4 Giorgio Buccellati

parallel with a modern situation is found in Polomé 1981, concerning the influence on the phonology of hitherto unwritten languages of “educated” transcriptions introduced by non-native speakers.)

(B) A second major filter affects Amorite. It seems certain that Amorite was never written down as such, i.e., there was no Amorite scribal tradition and accordingly, no Amorite texts. We only have Amorite personal names (plus a few technical terms), which were written down by Sumerian and Akkadian scribes, who developed their own conventions for rendering in writing the sounds of a language they did not normally speak, though presumably they did understand it (the existence of Amorite “interpreters” is presumably applicable only to the Sumerian south, at a time when Amorite was still something of a novelty).

1.1.2. *The writing system*

Inasmuch as we are dealing with a dead language, considerations about phonology are fundamentally affected by our understanding of the cuneiform writing system. Here are some of the most significant.

(A) There is a heavy reliance on various LOGOGRAPHIC subsystems, which apply especially to certain nouns and numerals, and exhibit considerable differences depending on time periods and text types (in modern transliterations, logograms are generally rendered in small capitals with their Sumerian value, e.g., URU for *ālum* ‘city’). Logograms are of no value for phonological reconstruction—so much so that one might question whether they are graphemes at all (see § 1.1.4). This, however, does not seriously affect our overall understanding of phonology since there are sufficient syllabic correspondences of words as lexical items to compensate for the widespread use of logograms as textual items.

(B) The use of HISTORICAL WRITINGS is more problematic, and it affects especially our ability to determine the time period at which a certain phonological change first occurred. For instance, the loss of final short vowels begins to be attested towards the end of the second millennium, but vowels in this position continue to be marked in the writing all the way down to the end of the documentation. In such cases we may assume that the linguistic phenomenon (as distinct from its graphemic representation) became operative across the board when first attested—a conclusion which is confirmed, in the particular case mentioned, by the fact that a short vowel in word-final position, written to indicate a presumed case ending, is often incorrect by morphological standards (e.g. *ālu*, *āla*, or *āli* are used indiscriminately for

āl), which indicates that the proper vowel was no longer supplied by any active linguistic competence.

(C) ORTHOGRAPHIC CONVENTIONS correspond to certain phonological regularities which we must define inferentially. For instance, while there are cuneiform signs used to render the presence of glottal stop, they do not normally occur in word-initial position; thus the word for ‘city’ may be written with the two signs <a-lum>, but not with the signs *<’a-lum>. If we do nevertheless assume that Akkadian words did not begin with plain vowel (hence regularly /’ālum/, rather than */ālum/, see § 1.3.8) it is because (a) we recognize graphemic rules next to graphemic values (see § 1.1.4), (b) we have very few examples of sandhi across word boundary (e.g. *libbālim* for *libbi ’ālim* ‘heart of the city’), and (c) we postulate a fundamental similarity with other Semitic languages.

1.1.3. *Phonology and graphemics*

The theory and application of GRAPHEMICS¹ is fundamental for an understanding of Akkadian phonology. This point needs to be stressed here in a special way because linguists who approach the Mesopotamian documentation without a proper understanding of the writing system can easily be misled into deriving false conclusions from what are presumed to be safe data. (Conversely, philologists are found sometimes to pay lip service to linguistic jargon, in that terms like “grapheme” or “phoneme” are used simply because they are perceived as more sophisticated than traditional terms like “sign” or “sound”; the consequent lack of proper definition makes for confusion rather than clarity.) I adhere to a narrower understanding of graphemics than usual. What phonemics is to meaning, graphemics is to sound. In other words, graphemics is the systemic correlation between graphic symbols and phonemes. Hence it is neither paleography (which deals with the shape of the symbols) nor orthography (which deals with the stylistic choice of alternative symbols from the same repertory). If every phoneme were represented by a single graphic symbol, the graphemic system would simply be a graphic overlay of the phonemic inventory. The alphabet (first invented in northeastern Syria towards the end of the second millennium—though some would push it back to the end of the third, Mendenhall 1985) comes close to this ideal, and it is, in this respect, a profoundly abstract linguistic accom-

1. What follows is based on my understanding of graphemic theory, as articulated in Buccellati 1979, 1982, 1984, 1990a. For an antithetic view, which calls into question the very notion of graphemics, see Daniels 1993.

plishment, besides serving as a very significant socioeconomic innovation. But unfortunately the cuneiform graphemic system is not quite as transparent. It is, in fact, a much more complex organism which includes multiple values and clustering laws.

Except for logograms, numerals written as digits, and semantic indicators (determinatives), all cuneiform signs are syllabic in nature. Each syllabic sign can render a variety of GRAPHEMIC VALUES, only some of which are made explicit in our modern system of sign-by-sign transliteration. For instance, in Old Babylonian any sign corresponding to a syllable with initial voiced stop may stand for a syllable with a homologous emphatic—i.e. the sign DA can stand for either /da/ or /ṭa/, though not for /ta/. GRAPHEMIC LAWS refer essentially to the clustering of values (hence they may be considered grapho-tactic in nature); for example, consonantal length may or may not be shown overtly, or else aleph may be marked by a simple vowel sign in word-initial position (e.g., <a-lum> for /ālum/) or in medial position after a sign ending in a consonant (e.g., <iš-al> for /iṣ'al/), but not necessarily after a sign ending in a vowel (e.g., <ba-a-bum> for /bābum/, not */ba'abum/, but <ša-a-lum> for /ša'ālum/).

In the American tradition of Assyriology (following I. J. Gelb), one uses the term “transliteration” to refer to the sign-by-sign rendering of graphemes (e.g., *i-il-la-ak*), and “transcription” to refer to the normalized rendering of their phonemic realization (e.g., *illak*).

1.1.4. *Graphemic analysis*

It is important to stress that not all writing systems, or subsystems, are equally graphemic in their import. In other words, not every written sign is a grapheme, nor is every grapheme fully explicit in terms of its correlation to phonemics. Ideograms, for instance, are not graphemic, and logograms are only minimally graphemic. Thus Akkadian numerals are for the most part written as digits: were it not for the very few instances of syllabic writing and for comparative inferences, these digits would be pure ideograms which would only convey the notion, but not the word, of the pertinent numerals (e.g., the notion of ‘one’ but not the word *iṣtēn*); and it is only to the extent that we can make the correlation between the digit 1 and the word *iṣtēn* that the sign acquires the status of a logogram. The graphemic value of a logogram is minimal, however, as it does not indicate, for instance, morphological variations, which must be interpolated from the context, through rules which are not properly graphemic, but rather grammatical. This is the problem that affects much of third-millennium cuneiform, especially

Eblaite. The reason is that the texts are set, by virtue of their content, within the Sumerian scribal tradition, and so logographic writing (based on Sumerian) is prevalent. Text categories which are less formulaic in nature (such as letters), and thus more likely to depart from logography, are unfortunately not well represented for the third millennium (and they are of course altogether missing for Amorite). Accordingly, the area which preserves the most explicit graphemic record for Old Akkadian is that of onomastics.

The Assyriological tradition has dealt all along with graphemic issues and has resulted in what must be recognized as an exemplary control of the data—the major achievement being the build-up of a cumulative list of sign values, culminating in von Soden 1948a (and von Soden and Röllig 1967). Gelb 1952 (2nd ed., 1961) and Huehnergard 1988 provide a model implementation of graphemics to individual dialects. But the development of a specific theoretical model for graphemic analysis is in its initial stages (Buccellati 1979, 1982, 1984, 1990a; Gaubelein 1976; Gelb 1961, 1970a; Kobayashi 1975; Lieberman 1977: 96–117; Platt 1993; Prosecky 1986; Reiner 1973a).

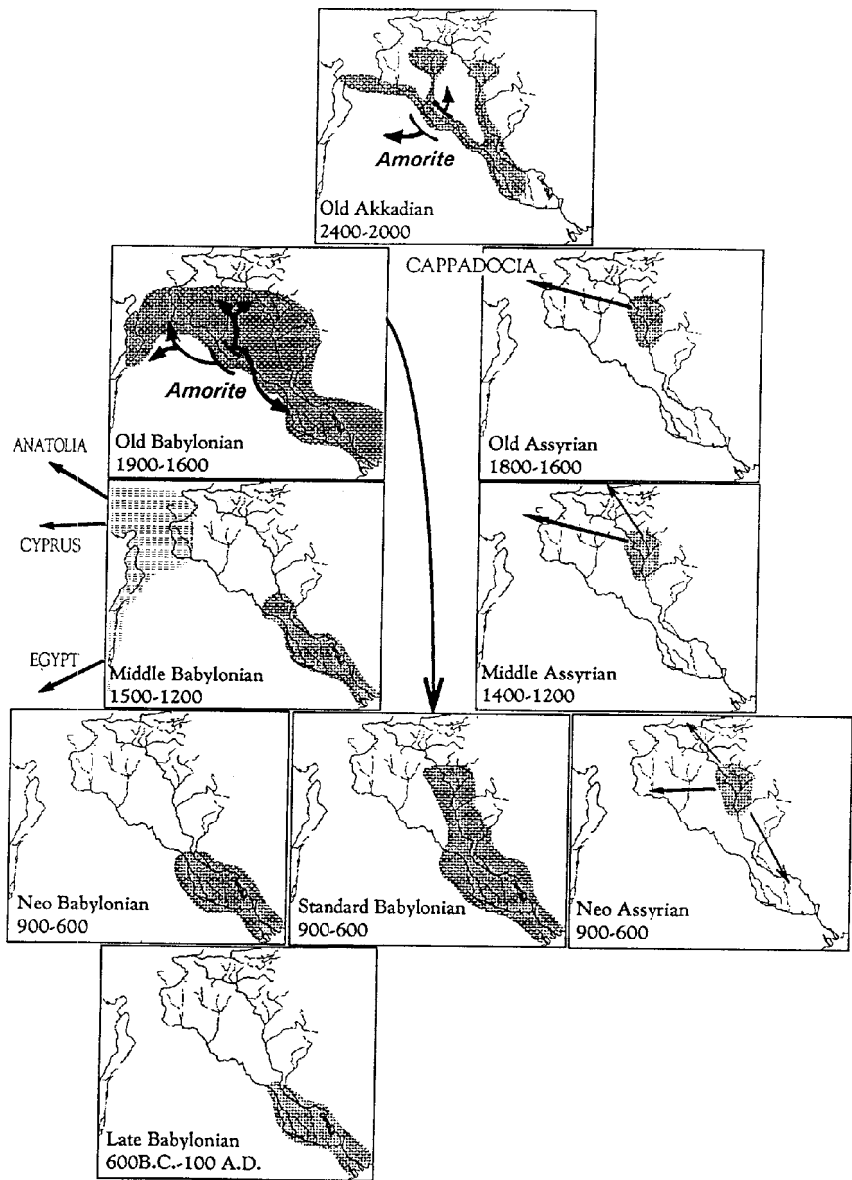
Two special considerations affect our understanding of Amorite graphemics. (1) Given our general assumption that there were no proper Amorite scribes (at least, no real Amorite scribal tradition, see § 1.2.3), the rendering of Amorite phonemes is to be understood through Akkadian graphemic practice. (2) Even though Amorite is typologically contemporary with Old Akkadian, for the most part it is preserved through the graphemic understanding of the later scribes; hence we must reckon with three different systems: third-millennium phonemes rendered by third-millennium graphemes (Old Akkadian); third-millennium phonemes surviving into the early second and rendered by second-millennium graphemes; second-millennium phonemes rendered by second-millennium graphemes.

1.2. Akkadian and Amorite

1.2.1. *Geographical and chronological distribution*

Akkadian is known from cuneiform texts dating from about 2500 B.C. to about the time of Christ, and Amorite from personal names (some 4000 text occurrences) preserved in cuneiform texts dating from about 2400 B.C. to about 1600 B.C.

In the third and early second millennium there is a substantial linguistic, and cultural, unity across the regions that correspond to modern-day Iraq and Syria: to this world we give the name Syro-Mesopotamia (see Map 1 for



Map 1. Geo-Chronological Distribution of Akkadian Dialects and of Amorite

a diagrammatic rendering of the dialects across this area). Akkadian (with Eblaite) is the language common to all city-states in this region; in other words, we are assuming the presence of native speakers of Akkadian from the Gulf to the Mediterranean. The term “Akkadian,” which continues to be used for the language as a whole, is taken from the name of the political entity which controlled most of Syro-Mesopotamia in the third quarter of the third millennium. Amorite is originally the language of the rural classes of the Middle Euphrates (see § 1.2.2), which extended originally to the steppe and began towards the end of the third millennium to migrate towards southern Mesopotamia. The term “Amorite” is derived from the Hebrew rendering of an original Amorite term, which is rendered as *Mardu* in Sumerian and *Amurru* in Akkadian, by which these people identified themselves as an ethnic group.

In the early second millennium there begins a marked differentiation between Babylonian and Assyrian, an innovation which mirrors the political development which resulted in the establishment of vast macro-regional states. (Obviously, the names of the two languages are derived from those of the political capitals.) Babylonian covers the entire area of Syro-Mesopotamia, while Assyrian is limited to a small enclave in the northeast (at least in terms of native speakers, excluding merchant colonies which are established as far west as central Anatolia). Individuals bearing Amorite names are now found over the whole of Syro-Mesopotamia, and we may assume that they were becoming more and more assimilated in the urban (Akkadian) setting; we presume that their competence as native speakers decreased as they became assimilated, until we lose all trace of even their names by the middle of the second millennium. (The political entity known as the kingdom of Amurru in the later second millennium may in fact be an efflorescence of earlier Amorite migrations, but it has no direct linguistic connection with Amorite.)

In the latter part of the second millennium, the geographical domain where we can expect native speakers becomes progressively restricted to the core areas corresponding to present-day southern and northern Iraq, respectively; Babylonian and Assyrian spoken (or often just written) outside these areas are juxtaposed to distinct local languages. (Of these, the most important is Aramaic, which may be seen as analogous to Amorite in its development from a rural base, except that it was protected by independent political institutions and it acquired the status of scribal language—hence it was preserved through textual evidence and not only through onomastics.) The

peripheral dialects, especially western Akkadian as found in the letters written in the local Syro-Palestinian courts of the late second millennium, are more properly to be considered as “bastardized” scribal adaptations than as genuine linguistic developments. Akkadian (in its Middle Babylonian variety) is used as the lingua franca for diplomatic exchanges among the contemporary world powers, including Egyptians and Hittites, but there is no evidence that it was ever used as a native language there.

From circumstantial evidence, it would appear that beginning at least with the 7th century Akkadian was no longer spoken, even though it continued to flourish in its written form all the way down to Seleucid times. The dialects of the first millennium come to be more and more closely associated with the scribal setting of the school than with the competence of native speakers. The “dialect” known as Standard Babylonian is specifically a literary hybrid, which harks back with a certain degree of almost archaeological awareness to Old Babylonian perceived as a model. The “dialect” known as Late Babylonian refers to the language of the periods after the loss of political autonomy (i.e., after the fall of Babylon), when Aramaic had become the prevailing language, under the political tutelage of the Persians, the Greeks, and the Parthians.

Comparative Semitics tends to consider Akkadian as a monolith, and to ignore Amorite altogether. The discovery of the archives of Ebla has redirected attention to the third millennium, yet often within the wrong perspective. Eblaite is a third-millennium language, while Akkadian continues to be seen as a second-millennium one and Amorite continues to be ignored. I think that one should instead take bolder steps in redrawing our perceptual map of early Semitic. Old Akkadian–Eblaite (or urban third-millennium Semitic) and Amorite (or rural third-millennium Semitic) should be studied on their own as reflecting the earliest documented phase of Semitic. Geographically speaking, they overlap with each other over the entire Syro-Mesopotamian area. Later Akkadian (though partly overlapping, in terms of the scribal documentation, with Amorite) is essentially a second-millennium language, the native base of which becomes progressively restricted to Babylonia and Assyria. The Akkadian of the first millennium is much less significant in terms of genuine linguistic development, because it is questionable how closely identified it might have been with the competence of native speakers.

1.2.2. Sociolinguistic position of Amorite

Amorite is universally regarded as a WEST SEMITIC LANGUAGE, and this is understood in a specific geo-historical sense. It is assumed, in other words, that

the carriers of this language were nomads who came to Mesopotamia from the Syrian steppe. This is not the place to dwell on the socio-historical details as to the nature of this “nomadism,” described in the literature through such categories as enclosed, full, or semi-nomadism. What matters, in our present context, is the overwhelming consensus that Amorite is West Semitic not just typologically, but also geographically and sociolinguistically. From this perspective, it would have to be treated as a separate chapter.

If I choose, instead, to present Amorite together with Akkadian² it is because I view Amorite and Akkadian as two closely related *SOCIOLECTS*, i.e., Amorite as the rural, and Akkadian as the urban, Semitic dialect of Syro-Mesopotamia in the third millennium (Buccellati 1992). This view is supported by socio-historical considerations which are not directly pertinent to our present discussion. But one should consider at least the following points. There is no indication of any movement from the West towards Mesopotamia, but only from the middle Euphrates, which is geographically at the very center of Syro-Mesopotamia. The typological similarity with later West Semitic may be explained by assuming a derivation of the latter, Aramaic in particular, from Amorite (Zadok 1993; but see Knudsen 1991: 883). The sequence would then be as follows. Amorite and Old Akkadian are the rural and urban counterparts of the same branch of Semitic in the third millennium, with Akkadian being, predictably, more innovative on account of the social context of its speakers. In second-millennium Mesopotamia, Akkadian continues its evolution independently. Amorite influence is practically nonexistent, since most probably all individuals with Amorite names spoke or at least understood Akkadian, and there were no new urban centers established by Amorites. In western Syria, on the other hand, Amorite speakers retained and developed their autonomous characteristics, especially following the fall of the kingdom of Khana (with its capitals Mari and Terqa, Buccellati 1990), which resulted in an urban vacuum on the Middle Euphrates and thus in the loss of a direct urban base for the Amorite speakers.

Regardless of how one may wish to view this overall reconstruction, the description of Amorite phonology given below stands on its own. The significance of Amorite in general has not been sufficiently appreciated for the study of early Semitic, and it is hoped that the treatment given here may help

2. The question has been posed by Garbini (1972: 151–54). His treatment is, however, different from the one proposed here.

to correct this situation. An important reason is the archaic nature of Amorite which places it, in my view, at an earlier stage of development than either Old Akkadian or Eblaite. Garbini (as early as 1960: 175–77) is one of the scholars who have emphasized the importance of Amorite, though he views it as being more innovative than I do.

The considerations I have adduced above are in support of my choice to consider Amorite together with Akkadian. If one objects to my understanding of linguistic correlations, one may simply wish to raise the discussion of Amorite to the status of a separate chapter. It should be noted in this connection that such distinctions are often heavily influenced by extra-linguistic considerations. Thus, Amorite is ignored because there is no literary tradition associated with it. On the other hand, an important reason why Eblaite is often considered a separate language (as in this volume), rather than an early dialect of Akkadian, is to be found in the geo-historical significance of the single archive which documents its existence; yet Eblaite and Old Akkadian are more closely linked to each other than, for example, Old Akkadian is to Neo-Babylonian.

1.2.3. *Phonological studies of Akkadian*

The year 1952 marked the beginning of a new era in the study of Akkadian grammar in general, and phonology in particular. It was then that both Gelb's *Grammar*³ and von Soden's *Grundriss* were published in their first editions. The latter is a monumental achievement, as to comprehensiveness of coverage (it includes all Akkadian dialects) and degree of philological control; methodologically it remains within a generic Neogrammarian mold. The former brings a whole new degree of theoretical sophistication to a proper linguistic study of Akkadian; in particular, its treatment of graphemics as the foundation for a proper understanding of phonology remains unmatched. Von Soden's grammar was to have a profound influence in the field, so that every other grammar that followed (with the exception of Reiner 1966; Huehnergard 1988; Izre'el 1991) was written strictly within the framework he had established (Aro 1955; Finet 1956; De Meyer 1962;

3. A very sharp criticism of Gelb's phonological reconstruction, especially with regard to the Old Akkadian sibilants, is found in Garbini 1972: 147–51. His objections concern especially the nature of graphemic correspondences and of the pronominal system. While recognizing the merit of some of his criticism, particularly with regard to broader methodological issues, a fuller philological analysis than Garbini's seems to me to ultimately validate Gelb's position, which is the one I essentially follow here.

Salonen 1962; Jucquois 1966; Hecker 1968; Giakumakis 1970; Wilhelm 1970; Mayer 1971).

Of fundamental importance for phonology is the study of the cuneiform syllabary, of which von Soden and Röllig 1967 represents the culmination. But it almost looks as though the two fields (grammar and syllabary) are conceived apart from each other, so that the full implications for phonology are not adequately articulated. Note how the major work on Neo-Assyrian “*Lautlehre*” (Deller 1959) is in fact primarily concerned with orthography rather than with phonology. It is in this respect that Gelb’s contribution stands out as a major methodological undertaking. In his work more than in any other scholar’s the accurate study of the syllabary is merged with a sensitive understanding of phonology, resulting in the most exemplary treatment of this topic.

Several studies on Eblaite (see elsewhere in this volume) touch on questions pertaining to Old Akkadian, but an in-depth comprehensive new study of Old Akkadian phonology remains a desideratum, especially considering that for Old Akkadian itself the documentary database has increased since Gelb’s groundbreaking work. Diakonoff 1985, 1991–92 and Faber 1985 represent a major step in this direction in terms of linguistic analysis, but they do not attempt to provide any systematic documentation.

In terms of a general treatment of phonology, a special place must be accorded Reiner 1966 and Diakonoff 1991–92. Reiner’s remains to date the most sophisticated linguistic analysis of Akkadian phonology; in particular it should be noted that she presented the first, and so far only, generative treatment of Akkadian phonology. Diakonoff is far-reaching in his implications, though his very strength (the correlation of Akkadian with Afro-Asiatic) limits the range and effectiveness of his Assyriological documentation and often obscures his elaboration of Akkadian phenomena.

1.2.4. Phonological studies of Amorite

The first modern treatment of Amorite phonology is Gelb 1958. Ironically, this remains also the only major work to provide a systematic treatment of the subject. Garbini has dealt with problems of Amorite phonology (1960: 19–80, 175–77; 1965; 1972: 23–96, 141–54), but from the point of view of specific comparative concerns. Neither Buccellati 1966 nor Huffmon 1965 covers phonology in any systematic way. While the last work by Gelb on Amorite (1980) is based on a rigorous understanding of phonology and provides all the essential data for a full discussion of the topic, it does not

articulate any of the details, for which the reader is referred to a forthcoming volume on Amorite Grammar that unfortunately Gelb did not live to complete; still, the volume is extremely significant not only because of the exhaustive database it provides, but also because important phonemic decisions are presented, particularly with regard to the establishment of the phonemic inventory (especially pp. 8f., 538f.).

Knudsen (1991; see also 1982a: 4–7) is of great import in that it proposes explicit criteria for assessing the relative degree to which Amorite consonants may be attributed full phonemic status (pp. 874f.), and divides the consonantal phonemes into three classes ranked in descending order of probability; unfortunately, Knudsen's contribution is only a summary which, by virtue of its self-imposed limitations, can neither develop a full articulation of its own argument, nor provide an adequate documentation for its conclusions.

1.3. Phonemics

1.3.1. *Phonemic inventory*

That we are able to reconstruct a plausible phonemic inventory for these languages is due to the essentially graphemic nature of the cuneiform writing system. For the most part, the scribal tradition was sensitive to the phonemic, rather than phonetic, dimension of Akkadian (Greenstein 1980; but see Diakonoff 1991–92: 3) and, though perhaps to a lesser extent, of Amorite. (It is due to this sensitivity that the invention of the alphabet was eventually possible.) The oppositions that are marked in the graphemic system correlate well to what we otherwise know, on the basis of living Semitic languages, about phonemic oppositions. It is for this reason that we can presume to draw up a phonemic inventory for dead languages on the basis of articulatory categories, a procedure which may appear at first bizarre considering that we have no record at all, acoustic(!) or descriptive, about the nature of such articulation. The exact articulatory nature of a group of consonants considered bilabial may well remain hypothetical, but its structural contrast with another category (which we will call, say, also hypothetically, dental) is beyond doubt. A simplistic way of expressing the net result of this procedure is to say that, were we to meet a living Akkadian-speaking informant and were we to try to speak Akkadian to him or her, we might be deemed to have an accent, but would not be incomprehensible, for the most part.

For Amorite, there is the additional filter of non-Amorite scribal transmission, which is especially critical in the rendering of phonological values. But just as Old Akkadian graphemics (though not paleography) was accurately preserved by Old Babylonian scribes, so it is to the credit of Mesopotamian scribes that they preserved many of the Amorite phonemic oppositions, even when they did not introduce any new graphic symbols.

One problem should be pointed out. In terms of their structural oppositions, the accuracy of the presumed articulatory identifications is not crucial. Such phonetic labels may serve just as such, i.e., as tags which approximate phonetic reality and are more convenient than non-descriptive labels (e.g., “labial” instead of “category A”). As with the reconstruction of proto-languages, emphasis tends to be on phonemic rather than specific phonetic identifications. Accordingly, we must be careful about raising articulatory terms, used as descriptive labels, to the status of real phonetic phenomena without the benefit of thorough critical analysis. To have done the latter is the merit of Diakonoff (1981; 1991–92: 1–4). A particular warning should be voiced against charting historical processes of change on the basis of definitions that have been introduced as labels. Change presupposes real sounds, not just labels. The case has been made recently for Proto-Semitic (Bomhard 1988), but it obtains of course for Akkadian and Amorite as well.

There are considerable difficulties in establishing a complete phonemic inventory for the third-millennium languages, and for second-millennium Amorite. This is due in part to the nature of the graphemic documentation, as mentioned above. But, considering the sizable amount of textual data available, the difficulty may also be attributed to the lack of in-depth studies on the subject. This is all the more remarkable in view of the widespread attention that has been lavished in recent years on the language evidenced by the texts of Ebla. It would seem that a proper understanding of Ebla phonology would be greatly enhanced by a detailed study of Amorite and Old Akkadian. This remains to be accomplished.

1.3.2. Consonants and semivowels

Besides giving an overall synopsis of the phonemic inventory (Table 1-1), I will deal with phonemic classes which present special problems and which have been the object of important recent contributions, especially the Old Akkadian sibilants.⁴ For comparative purposes I will refer to proto-forms or

4. This term is retained here because of its widespread use even though it describes improperly the acoustic, rather than articulatory, nature of the phoneme.

Table 1-1. Synopsis of Consonants and Glides (semivowels)
for Amorite, Old Akkadian, and Later Akkadian Dialects (OB+)

			Labial	Dental	Interdental	Denti- alveolar	Alveolar	Palato- alveolar	Velar	Pharyngeal	Laryngeal
Stops	Unvoiced	Am	p	t					k	ʕ	ʕ
		OAkk	p	t					k	ʕ	ʕ
		OB+	p	t					k		ʕ
	Voiced	Am	b	d					g		
		OAkk	b	d					g		
		OB+	b	d					g		
	Emphatic	Am		ʈ					q		
		OAkk		ʈ					q		
		OB+		ʈ					q		
Fricatives	Unvoiced	Am		s		ʃ	ʃ	ħ?		ħ	h
		OAkk		s	t		ʃ	ħ			
		OB+		s			ʃ	ħ			
	Voiced	Am		z	d						
		OAkk		z			[ž]				
		OB+		z							
	Emphatic	Am		ʂ							
		OAkk		ʂ							
		OB+		ʂ							
Sonants	Nasal	Am	m	n							
		OAkk	m	n							
		OB+	m	n							
	Lateral	Am					l				
		OAkk					l				
		OB+					l				
	Trill	Am					r				
		OAkk					r				
		OB+					r				
	Glide	Am	w					y			
		OAkk	w					y			
		OB+	w					y			

to Proto-Semitic, but I use these terms cautiously. We must remember that the validity of these forms is only proportional to the comparative basis from which they are derived, and that the posited proto-forms from which a historical derivation is assumed are in the first place projected back through a logical process. A proto-form is actuarial, as it were, rather than actual; it is an index for a set of correspondences, and should strictly be considered as such.

The inventory given here is in the form of a synopsis that offers a complete list for the consonants from each of the three major language groups. The listing should not be understood as describing historical development; the superposition of phonemes within the same cell is rather to be viewed as some sort of three-dimensional array which simply describes the phonemic attestations for the same type of articulation.

1.3.3. Special problems concerning Amorite

In the consonantal inventory of Amorite given here I use the list found in Gelb 1980: 9, which corresponds to the first two classes of Knudsen 1991: 874 (“established positive consonants,” i.e., consonants whose phonemic status is based on both unequivocal comparative and graphemic considerations, and “established neutral consonants,” based only on unequivocal comparative considerations, without graphemic support).

A questionable phoneme in Amorite is the unvoiced fricative palato-alveolar *ħ*. While it is attributed specific phonemic status in Gelb 1980: 8, 538, most of the entries given to support its existence are followed by a question mark. In Knudsen 1991: 874, *ħ* is considered a “non-established neutral” phoneme (i.e., a phoneme for which neither comparative nor graphemic criteria can be applied in a unequivocal way). I include it here in the inventory, but with a question mark.

Three additional consonantal phonemes have been proposed for Amorite,⁵ namely *ɖ*, *ʒ*,⁶ and *ǧ*.⁷ They are best considered, however, as historical reconstructions since there is no real evidence for their independent phonemic status. Gelb 1980: 8, 538 does not include them in his inventory, and Knudsen 1991: 874 defines them correctly as “non-established neutral” consonants. Accordingly, they are omitted in the inventory.

5. Gelb 1958 § 2.7.2; Knudsen 1991: 874.

6. Knudsen 1991: 874 uses the symbol *ʒ* for this phoneme; but this must be a typographical error for *ʒ*, which is analogous to Gelb’s *z*. The example *ṭabū* ‘gazelle’ (written *ṭabū* in Knudsen 1991: 874) corresponds to graphemic *ša-bu-um*, see Knudsen 1982a: 15.

7. Knudsen uses the symbol *ɣ* to render this phoneme.

1.3.4. *Laryngeals and pharyngeals*

Laryngeals are distinguished graphemically in Amorite but not in Akkadian, where *h* is subsumed under *ʾ*. The pharyngeals are distinguished graphemically as a set, but it cannot be determined whether they are further distinguished from each other. In Amorite, they are rendered by signs with *ḥ*. In Akkadian there is no overt graphemic marker, but we can infer that they were still present in Old Akkadian, because of the way in which they affect the vocalism of the word in which they appear: in Old Akkadian the vowel in contact is *a*, except in closed syllables beginning with pharyngeal, where it becomes *e*,⁸ while in the later dialects all *a* vowels in the core of the word become for the most part *e* (see § 1.4.2). Table 1-2 lists the correspondences for laryngeals and pharyngeals among the various dialects and their posited Proto-Semitic equivalent.

Table 1-2. Correspondences among Laryngeals and Pharyngeals

*	Am	OAkk	OB+	Am	OAkk	OB+	Gloss
ʾ	ʾ A	ʾ A	ʾ A	<i>ʾabu</i>	<i>ʾabu</i>	<i>ʾabu</i>	‘father’
h	ḥ ḤA	ʾ A	ʾ E	<i>haddu</i>	<i>ʾadad</i>	<i>ʾadad</i>	‘storm god’
‘	‘ ḤA	‘ A	ʾ E	<i>ʾzb</i>	<i>ʾzb</i>	<i>ʾzb^a</i>	‘to leave’
ḥ	ḥ ḤA	‘ A	ʾ E	<i>bḥr</i>	<i>bʾr</i>	<i>bʾ^b</i>	‘to choose’
ḡ	? ḤA	‘ A	ʾ E	?	<i>ʾrb</i>	<i>ʾrb^c</i>	‘to enter’

a. Babylonian *ezēbu*, Assyrian *ezābu*.

b. Babylonian *bēru*, Assyrian *beʾāru*.

c. Babylonian *erēbu*, Assyrian *erābu*. Knudsen 1991: 874 gives *purgūšu* ‘flea’ as an Amorite example for *ḡ*. See Diakonoff 1985: 20 for an Eblaite equivalent.

1.3.5. *Sibilants*

The treatment of the sibilants (see notes 4 and 5) presents us with severe problems. (A) The graphemic rendering, while relatively consistent within each dialect, appears to us confusing when comparing different dialects. (B) The phonetic realization of some of the phonemes is in part uncertain. (C) The notations used in the literature are often ambiguous.

It is regrettable that Gelb, while distinguishing clearly the various categories on the theoretical level, does not carry this over to either the text of his *Grammar* (1961) or the entries of his *Glossary* (1957); rather, he uses capital *Š* to subsume without differentiation *š*, *ś*, *ž*, and *ṣ*. In their edition of Old

8. For a fuller statement of the pertinent rules see Gelb 1961: 123–25.

Akkadian royal inscriptions, on the other hand, Gelb and Kienast (1990) use a hybrid system whereby the phoneme *š* is marked in the transliteration (e.g., *u-ša-am-qí-it*), even though the value is not recognized in the standard syllabaries. Von Soden (1965–81) uses the same symbol *š* to render both **š* (he writes *šīāmu* s.v. *šīāmu* ‘to set’) and **ṣ* (*wašābu* s.v. *wašābu* ‘to dwell’).

Diakonoff 1985, 1991–92 and Faber 1985 present a divergent interpretation whereby they assume an affricate realization for the sibilants (see already Steiner 1977: 144–48, 159; 1982: 70–74). Their theory is of great interest and is all the more noteworthy since they arrived at it independently of each other (Diakonoff 1991–92: 55, n. 61) and from different points of departure. While it affects especially phonetic realization⁹ rather than phonemic distribution, it is important in that it proposes new possibilities for historical derivation and for morphophonemic alternations, though not without difficulties (see § 1.4.3). I will not follow their theory here because I feel that more reflection is needed before their results may be accepted. As already noted in the case of Gelb, it is regrettable that Diakonoff himself does not carry through his notation consistently: generally he uses the traditional notation *š* (e.g., *šin-a* ‘two’, 1991–92: 17), and only occasionally the notation he proposes (e.g., *sawxat* for traditional *šamḫat*, 1991–92: 114; *uṣ:īṣ* for traditional *uššiš*, 1991–92: 52).

Here I will attempt to clarify the situation by comparing various aspects of the problem. Table 1-3 summarizes the terms of the problem with regard to graphemic rendering and divergent modern notations. Graphemes are grouped in sets which share the same consonantal realization, and have a different vowel for each sign. (It should be noted that accent marks and subscripts have no phonemic implication, but simply serve as standard Assyriological indices for the identification of homophonous cuneiform signs.) For each graphemic set I show the variant modern notations used to render the phonemic value of the consonant in question for different dialects and periods. It must be stressed that this table establishes only a synopsis of graphemic correspondences; in other words, the chart should not be understood to say that Old Akkadian *š* corresponds to Old Babylonian *s*, but only that graphemes of the class *sÁ* are used to render *š* in Old Akkadian and *s* in Old Babylonian.

Table 1-4 singles out those phonemes which present divergent realizations in the different dialects and periods, and it shows their correspondence with

9. Diakonoff 1991–92: 39–41 goes into great detail as to the precise phonetic “pronunciation” of this series.

Table 1-3. Graphemic Sets and Modern Phonemic Notations for Sibilants with Divergent Realizations^a

GRAPHEMES	AMORITE			OAKK					OB+	
	G	K	B	G	GK	D	F	B	c	D
SA SE ₁₁ SI SU	ś	ś	š	{ š, š ₁ ś, š ₂ }	š	c	s	š	s	c
SÁ SU ₄ ŠÈ	ž	<u>d</u>	<u>d</u>	ž, š ₄	š	c	s	š	s	c
ŠA ŠI ŠU	š	š	š	<u>t</u> , š ₃	ś	č	<u>t</u>	<u>t</u>	š	{ ^s č
ZA ZI ZU	s	s	s	s	s	c	ts ^b	s	(s)	c
	{ š	š	š	š	š	č	ts'	š	š	č
	z	z	z	z	z	ž	ds	z	z	ž

a. B, Buccellati; c, common use; D, Diakonoff; F, Faber; G, Gelb; GK, Gelb–Kienast; K, Knudsen.
b. Digraphs of the type *ts* stand for single (affricate) phonemes; similarly, *ts'* stands for a single phoneme which Faber defines as a glottalic pressure affricate. Note that Faber 1985: 105 considers *ts*, *ts'*, and *ds* to be phonetic realizations of one and the same phoneme *z*.

Table 1-4. Phonemic and Graphemic Correspondences for Sibilants with Divergent Realizations

*	Am	OAKk	OB+	Ar	Am	OAKk	OB+	Ar	Gloss
š	}	š	SA	š	š	š	š	š	'name'
ś		ś	SA	ś	šym	šym	šym	(šym)	'to place'
<u>t</u>	š	šA	<u>t</u>	šA	<u>t</u>	yšb	wšb	wšb	'to dwell'
<u>d</u>	<u>d</u>	SÁ	z	ZA	z	ž	ž	ž	'to take'

regard to a posited proto-form, and, for ease of reference, with Arabic (or South Semitic) as well. I give here only the phonemic notation that I have chosen to use, plus the graphemic rendering with only one sign for each of the sets given in Table 1-3. I also add representative word examples.
Following are comments on each of the phonemes listed in Table 1-4. I relate in some detail the definitions given by various scholars, because they are often presented cryptically in the literature, and it is difficult to correlate opinions about what should be simple facts.

š is the symbol for a lateral that is reconstructed for Proto-Semitic,¹⁰ but is not preserved as such in either Amorite or Akkadian.¹¹ Even though it is used often in the Assyriological literature, this notation should be avoided when dealing with those two languages.

š is a symbol I am using to render what I assume to be a distinctive Amorite phoneme, at least in terms of its derivation. The Amorite phoneme corresponds to “Proto-Semitic” *š and *ś. Though it is often transliterated as ś, it does not appear to be a lateral,¹² because of the writing with signs of the class SA. It seems possible to postulate a phonetic development similar to Arabic, i.e., a change in the direction of s,¹³ with place of articulation shifting toward the dental position. But that it could not simply have merged with s is also indicated by graphemic considerations.¹⁴ Hence I am postulating, on a purely indicative basis, that this phoneme may have been realized as a denti-alveolar fricative. At any rate, the phoneme is distinct from any other in terms of its correspondences (i.e., it corresponds to *š and *ś, but not to *t), and this by itself is sufficient to justify the use of a different symbol.

t is the symbol used for the voiceless interdental fricative (θ) in the posited proto-form. The Old Akkadian correspondence is generally transcribed as ś in the literature, but, since there is no reason to assume a lateral realization for this phoneme in Old Akkadian, I prefer to retain consistently the notation t on the assumption that the original interdental realization was preserved.¹⁵

10. The autonomy of this phoneme for “Proto-Semitic” is controversial, but it can at least be safely postulated as an antecedent to Amorite and Old Akkadian (Gelb 1961: 34f.).

11. Except possibly in Old Akkadian, see Reiner 1966: 110; Greenstein 1980, with review of previous literature; also Diakonoff 1985: 22, where the notation *q* is used. The strongest indication in its favor is based on the explanation of the assimilation š+š > ss as presupposing a second consonant ś, see § 1.4.3, where I also give reasons why it seems nevertheless better not to include it in the standard phonemic inventory.

12. As Greenfield 1969: 94 seems to suggest, on the assumption that *š coalesced with *ś, so that both came to pronounced ś.

13. Knudsen 1982a: 5 says that its phonetic realization is “similar to Old Babylonian s.” Gelb does not address the issue of articulation for this phoneme.

14. To render š, signs of the SA class are used next to signs of the šA class, and only signs of the Aš class are used. To render s, only signs of the SA and AS classes are used.

15. “Possible pronunciation in the direction of Arabic t, perhaps not in Mesopotamia proper but in an outlying region” (Gelb 1961: 37). In the chart in *ibid.* p. 39, the “sound” of this phoneme is indicated as t', which, according to the discussion on p. 33, would seem to stand for an aspirated correlative of t.

ʃ is the standard symbol for the voiceless palatal or alveo-palatal fricative. It continues in OAkK and OB+ (where it comes to subsume other phonemes as well). Amorite ʃ , on the other hand, corresponds to PS t , and is presumed to be phonetically the same as Akkadian ʃ ,¹⁶ because the graphemes used to render it are those used to render ʃ in contemporary Old Babylonian.

z is the standard symbol for the voiced denti-alveolar fricative. It corresponds in OAkK and OB+ to PS ð . In addition, z is also the normal Am, OAkK, and OB+ correspondence for PS z (not shown in the chart above).

ð is the voiced interdental, which has a correspondence in Amorite.¹⁷ Gelb also thinks that there may be an archaic ž in OAkK, derived from PS ð .¹⁸ Since this is, however, uncertain, I omit it from the phonemic inventory given above.

1.3.6. Vowels

Both Old Akkadian and later Akkadian share the same inventory of four vowels (Table 1-5).¹⁹

Table 1-5. Akkadian Vowels

	Front	Back
High	i	u
Low	e	a

16. “Probabilmente una sibilante ʃ piuttosto che una dentale fricativa t ” (Gelb 1958 § 2.7.9). “Similar to Old Babylonian s ” (Knudsen 1982a: 5); “one of the two Amorite phonemes [ʃ and ʒ] probably represented the palatal groove spirant ʃ ” (Knudsen 1982a: 6).

17. “Voiced interdental spirant” (Knudsen 1982a: 4; he uses the symbol d to render this phoneme). “Support in favor of [a pronunciation as ž] is to be found in the phonemic analysis of Old Akkadian ..., Amorite and Ugaritic” (Gelb 1980: 8b).

18. “Signs of the ʒ_4 class are to be considered as leftovers from a period in which Akkadian recognized a phoneme ž ($=\text{ʒ}_4$) < d ” (Gelb 1961: 38). In the chart in *ibid.* p. 39 the “sound” of this phoneme is indicated as t , i.e. the non-aspirated member of the pair discussed on p. 33. This phoneme is omitted from the list given on p. 119.

19. Diakonoff 1991–92: 68 has noted that in Akkadian primary nouns, the vocalic system essentially excludes u , and that e is used only in Sumerian loanwords, while conversely in verbal roots the vocalic system excludes e except as allophone of a (on the phonemic status of e see also p. 123). This interesting observation pertains more to the study of word formation than of phonology per se. On the phonemic status of e see Izre’el 1987.

Diakonoff (1991–92: 123–25), following a suggestion by Lieberman (1979), concludes that at least from OB on Akkadian had a phoneme *o* (and, with corresponding length, *ō*). He argues in part from the observation that the presence of a phoneme *e* causes the system to be asymmetrical, and therefore unstable. However, the phoneme *o* (and *ō*) is in fact understood by him as a phonetic realization of *u* (and *ū*), so that the asymmetry noted is simply shifted to another plane. Accordingly, I prefer to retain the traditional scheme as given above.

The situation in Amorite seems to be limited to three vowels (Table 1-6).

Table 1-6. Amorite Vowels

	Front	Back
High	i	u
Low	a	

The vowel *e* does appear in the writing of Amorite, but, given the absence of contrasting minimal pairs, it may more properly be understood as an allophone of *i* or *a* (Knudsen 1991: 870). Gelb 1958 § 2.1.1–4 assumes the existence of a phoneme *ē*, and Knudsen 1991 : 870 the existence of both *ē* and *ō*, i.e., long vowels without a corresponding short vowel. For neither vowel, however, can one adduce a convincing minimal pair. Thus *ē* and *ō* are considered here as special phonetic realizations deriving from the contraction of diphthongs.

1.3.7. Suprasegmentals

All consonants and vowels can be lengthened in Akkadian and Amorite. Considering certain practices which are current in standard Assyriological tradition, and which are misleading for a proper understanding of length, it bears mentioning that this phenomenon is to be understood as the holding of the articulation for a fraction of time. (1) One speaks of “doubling” of consonants (so much so that the derived verbal stem with lengthening of the middle radical is labeled with D for doubling): but there is obviously no reason to suppose that the articulation was repeated twice, only that it was held longer. (2) One uses a two-tiered notation for long vowels, e.g., *ā* and *â*, but there is no conclusive evidence that this corresponds to a phonemic distinction; in other words, there are no minimal pairs to show that there were two contrasting degrees of length. The distinction made in standard Assyriological notation between *ā* and *â* is etymological rather than

phonemic (circumflex is used to mark derivation from contraction). It also leads to confusing and contradictory applications, so that it should best be ignored.

It is possible that the phenomenon behind the dual notation of length envisaged by our modern grammatical tradition may, instead, have something to do with stress (on this see Sarauw 1939; Knudsen 1980; Greenstein 1984: 24–27; Diakonoff 1991–92: 104–15). Consider the following: *panū* ‘face’ vs. *pa’nū* ‘first’. If there is in fact a contrast based on stress (thus also Diakonoff 1991–92: 111), it is because the second word derives by contraction from *pan-ī-u*. The traditional view about Akkadian stress is that it falls on the first long syllable from the end of the word, except that morphemic length in final position is disregarded (as with *panū*). A possible case for an alternative theory has been made by Reiner (1966: 38, following a suggestion of Poebel 1939: 60): she thinks it probable that in Assyrian, at least, primary (non-phonemic) stress would fall on the first syllable of the word. Diakonoff (1991–92: 104–15) has argued convincingly in favor of the traditional view, which is retained here.

1.3.8. *Phonotactics*

A number of distributional limitations affect the actual cooccurrence of phonemes in a variety of ways. I will describe them here as functions of word boundaries, clustering, and syllabic structure.

1.3.8.1. *Word-initial position*

Any simple consonant may occur in word-initial position in both Amorite and Akkadian; of the semivowels, only *y* may occur in Amorite, and only *w*²⁰ in Old Akkadian and Old Babylonian/Old Assyrian (in the later Akkadian dialects initial *w* is also excluded). The exclusion of initial long consonant or consonantal cluster is no surprise, since it is a common Semitic feature. Only two additional comments are in order. (1) We assume that vowels are not allowed in word-initial position on account of comparative and (to some extent) graphemic considerations. The alternation in the writing of forms like <i-il-la-ak> and <il-la-ak> ‘he goes’ suggests that in the first instance ’ was overtly indicated (one could transliterate <’i-il-la-ak>), while in the second the notation was omitted because its presence was assumed as automatic. In modern transcription, ’ is regularly omitted, for the same reason (hence our writing *illak* really means *’illak*); see also § 1.1.2. (2) The con-

20. For the exclusion of initial *y* in Akkadian see Reiner 1964; Gelb 1970a: 536–43.

trast between ^ʕy in Amorite and ^ʕw in Akkadian is particularly significant because the Amorite situation is universally assumed to be an innovation that closely links it with West Semitic.

1.3.8.2. Word-final position

Any simple vowel (short or long), and any simple consonant, may occur in word-final position. Specifically, no semivowel, long consonant, or consonantal cluster may occur in this position. These rules apply equally to Amorite and Akkadian.

1.3.8.3. Clusters

Consonantal clusters of two, but not more, consonants occur in word-medial position. Vocalic clusters occur in Amorite, Old Akkadian, and Old Babylonian/Assyrian when a morphemic boundary intervenes between them, e.g., *rabī+at* ‘she is great’ (with but few exceptions, e.g., *’iqīaš* ‘he donates’, without morphemic boundary). It is likely that in these cases a glide was present (*rabīyat*, *’iqīyaš*). In some cases, both graphemics and word structure suggest the presence of a long glide, e.g., *dayyānum* ‘judge’. Vocalic clusters do not occur in later Akkadian.

1.3.8.4. Syllabic structure

(See especially Greenstein 1984 with the reviews Edzard 1986; Knudsen 1986.) The following distributional rules apply: the components of a cluster are always separated by syllabic boundary; long consonants are treated like a cluster, as if we had reduplicated consonants with syllable boundary between them, even though there is no reason to assume double articulation (see § 1.3.7); no syllable begins with a vowel, except as second element of a vocalic cluster; no long vowel occurs in front of either a consonantal cluster or long consonant. These rules apply equally to Amorite and Akkadian.

Vowel harmony is a very distinctive Assyrian phenomenon: *a* in a short medial syllable which follows a stressed syllable assumes the quality of the following vowel, e.g., OB *’iṣbatū* ~ OA *’iṣbutū*. (One may consider the possibility that in OB short *a* may have been used in writing to render *ə*, see § 1.3.9; if so, *iṣbatū* and *iṣbutū* may be understood as graphemic equivalents rendering one and the same form *iṣbətū*. In other words, vowel harmony would be a graphemic rather than a specifically phonemic feature, in the sense that both Babylonian signs with *a* and Assyrian signs with *a/e/i/u* would stand for a purely phonetic *ə*.)

Reiner's suggestion (1966 § 4.1.2.5) of an equivalence between V:C and VC: is applicable only in prosodic terms, and should not be taken to mean (as seems to be the case in Diakonoff 1991–92: 116) that no phonological opposition exists. An opposition is clearly apparent in such morphemically diverse minimal pairs as *šūma* 'he himself' vs. *šumma* 'if'; *dānum* 'judge' vs. *dannum* 'powerful'; *īkūnū* 'they stood firm' vs. *īkunnū* 'they stand firm'.

1.3.9. *Phonetic realizations*

While the cuneiform writing is essentially phonemic in nature, there are clues to abnormal phonetic realizations which, as far as we can tell, fall outside the phonemic range. Some of the more interesting pertain to vocalic quality (*ö, ü*, von Soden 1948b; *ə*, Buccellati forthcoming §§ 14.1, 55; see also § 1.3.6, and § 1.4.2 herein); nonemphatic realization of velar emphatic (*k* for *q*, Knudsen 1961); spirantization of stops (*p, b, t, d, k, g*, von Soden 1968); realization of *m* as *w* from OB on (Diakonoff 1991–92: 125); stress (Aro 1953). All these phenomena have been observed for the dialects later than Old Akkadian. For an affricate realization of the sibilants see § 1.3.5.

1.4. Phonological change

1.4.1. *Historical changes affecting individual phonemes*

The correspondences which have been noted above for laryngeals and pharyngeals (§ 1.3.4), sibilants (§ 1.3.5), and vowels (§ 1.3.6) are the most significant in terms of a presumed derivation from a common proto-form. Amorite and Old Akkadian are, in different ways, relatively close to the posited proto-forms.

Amorite appears to be the most archaic. It preserves (1) the laryngeals and pharyngeals, presumably in their differentiated form; (2) *ə*; and (3) a restricted vocalism with only *i, a, u*. The major innovations in terms of the inventory are the change of **t̰* to *š*, and of *š* and *ś* to *s*'. The former is in common with later Akkadian and with other Semitic languages, while the latter seems to be peculiar to Amorite.

Old Akkadian occupies an intermediate position. It is more archaic than Amorite only in its preservation of *t̰*, while the other sibilants appear already in the same form as in later Akkadian. Both laryngeals appear as ' as in later Akkadian. The pharyngeals appear to have merged in a common consonant, for the quality of which we have no indication in the writing, but which we assume to be '.

By the beginning of the second millennium, Akkadian has undergone radical changes, in particular the reduction of Old Akkadian ' and ' to ' and ' to ';

the reduction of Old Akkadian *š* and *ṣ* to *š*. As a result, the later phonemic inventory of Akkadian, though still quite ancient in terms of chronology (about 1900 B.C.), is to be considered very recent in terms of typology.

1.4.2. Historical changes affecting phonotactics

A great variety of phonotactic phenomena can be identified over the long history of Akkadian. Here I will only mention a few that seem particularly characteristic of each major period.

The alternation between Akkadian *ʿw* and Amorite *ʿy* is generally explained as Akkadian conservatism and Amorite innovation, respectively, vis-à-vis a presumed proto-form *ʿw*.

Diphthongs (understood as vowel+semivowel in closed syllable) have been posited for proto-forms such as **mawtum* from which Akkadian *mūtum* would have derived. Such proto-forms, however, may reflect an undue emphasis on consonantal triradicalism; if we assume that length may function as a radical (Buccellati forthcoming § 39.1), then forms like *mūtum* may in fact be archaic, and those with *w* recent (introduced to fit a triconsonantal scheme).

The reduction of Old Akkadian *‘* (corresponding to *‘*, *h*, and *g*) to *’* was compensated for by a change in vocalic quality from *a* to *e*. This change is partly morphophonemic (see § 1.4.4), because the change occurs only within the core of a word that is not affected by external inflection; e.g., OB *‘ezēb-am* (not **‘ezēb-em*) ‘to abandon’ with accusative ending (contrast *‘alāk-am* ‘to go’, also with accusative ending). In Assyrian, only the vowel in contact with *’* shifts to *e*, i.e., *‘ezāb-am*, *‘iltanaqqē* ‘he repeatedly took’ (contrast OB *‘ilteneqqē*).

In Middle Babylonian the consonantal clusters formed by *š* plus a dental stop or an alveolar fricative, *š* shifts to *l*; e.g., OB *‘išṭakan* ~ MB *‘iltakan* ‘he then placed’.²¹

The loss of short vowel in final position is a phenomenon which is found in the major Northwest Semitic languages of the first millennium, with important consequences for morphology (case endings) and syntax (determina-

21. No phonetic explanation has been offered for this shift. Very tentatively, one may entertain the following conjecture: dental lateral *š* may have been preserved in early Aramaic dialects (for Akkadian, see Diakonoff 1985: 22), and the shift to a liquid lateral *l* may have started with clusters of the type *št* and then spread to *šr* and others. The evidence for *š* in Ancient Aramaic is controversial (Degen 1969 § 13; Segert 1975 §§ 3.2.3.5.3, 3.2.8.3–6), but note for instance how the Masoretic rendering *Kašdīm* (Hebrew) / *Kašdāy* (Aramaic) corresponds to later Babylonian *Kaldu*. For suggestions of *š* in Akkadian see § 1.3.5.

tion). It is interesting to note how Northwest Semitic syntax compensates through the introduction of the article, which was never introduced in Akkadian: this is one of the important indications that first-millennium changes within Akkadian, minimal as they are, no longer take place within the context of a living language.

1.4.3. *A note on morphophonemic alternations*

Many of the phonological changes which are discussed in standard treatments of Akkadian and Amorite phonology are in fact alternations conditioned by morphemic factors (Reiner 1966: 104–12). Since they presuppose an understanding of morphology, they cannot be taken up here, particularly because a fully coherent statement along those lines entails that the entire system of the so-called weak verbs be presented as part of morphophonemics, and this is clearly beyond our present scope (an extensive section is devoted to this topic in Buccellati forthcoming). I will only mention here four rules which are of particular interest because they apply only in Akkadian, and not in Amorite. (1) It is generally (but improperly) stated that in a sequence of three short vowels the middle vowel is dropped. The full conditions may be stated instead as follows: a word resulting, through internal inflection, in a sequence of three syllables of which the first two are short²² is realized with the elision of the vowel in the middle syllable; e.g., Akk. {damiq-um} /damqum/ ‘good’, contrast Am. *malak-a* ‘he has ruled’ (see Huffman 1965: 89).²³ (2) The pattern MAPRAS is realized in Akkadian with initial *n* if any of the radicals is a labial, whereas initial *m* is retained in Amorite, e.g., Akk. *našparum* ‘envoy’, contrast Am. *mašparum*. (3) Two emphatic radicals are found to cooccur in Amorite, whereas in the Akkadian correspondence one of the radicals is realized as unvoiced (Geers 1965); e.g., Am.

22. This formulation implies that at least the consonant of the third syllable be part of internal inflection, as in {damiq+um}. The rule so stated excludes (1) a case like {iṣbat+u+šu} (/iṣbatušu/, not /iṣbassu/) ‘that he seized him’, where only the sequence /bat/ properly belongs to internal inflection; (2) primary nouns, e.g., *išar+um* (not **išrum*) ‘normal’ (see Goetze 1946, 1947); (3) loanwords, e.g., *Uruk+ī+u* (not **Urkūm*) ‘Urukian’, *gabadiḫ+u* (not **gabdiḫbu*) ‘paraper’; sequences derived (4) through external inflection, e.g., *šarra+šunū* (not **šarrašnū*) ‘their king’ or (5) through onomastic name composition, e.g., *Abu-ṭāb* (not **Ab-ṭāb*) ‘Father-is-good’.

23. A related case is seen in the formation of certain feminines where Akk. opts for a plain -*t*- marker, whereas Am. uses -*at*-, e.g., Akk. *watar-t-um* ‘exceeding’, contrast Am. *yatar-at-um*. Occasional exceptions in Amorite, which are explained as orthographic in nature, are quoted by Knudsen 1991: 872f.

qsr; Akk. *kšr* ‘to bind’. All these statements are applicable in Old Akkadian as well as in late Akkadian.

(4) A case of particular interest is the Akkadian realization of {š+š} as /ss/ where the first š is part of a nominal or verbal base²⁴ and the second š stands for the first consonant of the pronominal suffixes of the third person. The dental realization is baffling, because it seems to imply a different phonetic status for suffixal š. It has in fact been suggested that suffixal š may have been a distinct phoneme ś, which would continue the ś of the PS pronominal suffix.²⁵ If so, the dental realization as /ss/ may be explained on the basis of an original cluster š+ś, e.g., *ēpuš-šunūšim* /epussunūšim/ ‘do it to them!’. However, this would be the only evidence for a phoneme ś in later Akkadian. Hence, rather than positing ś as a distinct phoneme, it is generally preferred, e.g., Reiner 1966: 110, to posit instead a special morphophonemic rule. This particular phenomenon is all the more remarkable since suffixation is very productive even with neologisms in later periods, e.g., Neo-Babylonian *išag-giš+šu* /išaggissu/ ‘he will kill him’ (von Soden 1965–81: 1126b).

It should be noted here that the affricate interpretation of the sibilants (see § 1.3.5) seems at first to account well for some analogous instances where a similar assimilation occurs. Thus, if, e.g., *māt+šu* ‘his land’ is interpreted as *māt+su*, the resulting phonetic realization would indeed be /mācu/. But all is not as simple as it seems at first. For instance, the affricate realization /mācu/ yields a short consonant, whereas graphemic considerations suggest that the consonant in question was long; how then would we explain the realization /maccu/ from /mat+su/, since [ts] is /c/ and not /cc/? More importantly, how does a cluster of two identical dental sibilants (e.g., *epus+sunūšim* in Diakonoff’s transcription for *epuš+šunūšim* in the standard transcription) yield a long affricate (/epuccunūšim/, again in Diakonoff’s transcription)?

24. This formulation excludes, for instance, *ašar-iš+šu* ‘to his place’, since the affix *-iš* is not a nominal or verbal base. The affix *-iš* may in fact go back to *-is* (Rabin 1969: 192). Such exclusion, too, casts doubt on an interpretation of suffixal š as ś, but see Reiner 1966: 110. Reiner reports as an additional exception the realization /šš/ when the final radical š derives from PS ś, as with the root *q:š* ‘to donate’. However, forms with such a realization (e.g., {iqīš+ši} /iqīšši/ ‘he donated her’) are late; OB forms generally show a realization /ss/ (e.g., {lūqīš+šim} /lūqissim/ ‘let me donate to her’). Note that long š is otherwise well attested in a variety of other environments, e.g., *kaš-kašš-u* ‘very powerful’ from the root *kšš*.

25. See above, n. 12. Note that no such environment may occur in Amorite given the lack of suffixes beginning with š.

1.4.4. *Free variation*

Alternations which cannot be explained with any degree of plausibility are best considered simply as free variants. Two examples may be mentioned here. (1) Vocalic alternations, e.g., Akkadian *ʾušakniš* ~ *ʾušekniš* ‘he subdued’. (2) Contraction across word boundary (sandhi), e.g., Akkadian *libbālim* ~ *libbi ʾālim* ‘within the city’, or Amorite *yarkibaddu* ~ *yarkib Haddu* ‘Adad rides’ (Knudsen 1980: 7f.).

1.4.5. *Historical development*

From the descriptive presentation of inventories and changes, two important points emerge. (A) Akkadian, already at the stage of Old Akkadian, presents more innovations than Amorite. Two concurrent explanations may account for this. On the one hand, Akkadian was subject to the direct influence of Sumerian—something which has long since been recognized. On the other hand, if I am correct in viewing Amorite as the rural counterpart of Akkadian (see § 1.2.3), we can easily accept the conservative nature of its documentation, even if chronologically later in terms of scribal attestation.

(B) Within Akkadian itself, the major changes occur at the level of phonotactics, whereas the inventory as such undergoes only relatively minor changes after Old Akkadian (and fewer changes yet if one accepts the reconstruction of the sibilants by Diakonoff 1985, see § 1.3.2). In particular, no substantial change can be identified within the first millennium. Again, two concurrent explanations may be given. The first is that, within the second millennium, the scribal tradition acted both as a filter in the transmission of the evidence and, to a more limited extent, as a brake to change. Phonotactic changes are the ones that are more difficult to hide with standard scribal mechanisms. The second explanation is more complex. Within the first millennium, the influence of Aramaic was so pervasive that, in effect, it hardly showed at all. By this I mean that it was a social rather than a linguistic influence: somewhere in the early centuries of the millennium, speakers were at best bilingual, and soon afterwards (probably by the eighth/seventh century) Aramaic took over as the only spoken language, while Akkadian continued merely as a literary medium. Hence it is that the linguistic influence on Akkadian is essentially limited to the lexical and syntactic spheres.

Bibliography

- Aro, Jussi. 1953. “Abnormal Plene Writings in Akkadian Texts.” *Studia Orientalia* 19/11. Helsinki: Societas Orientalis Fennica.

- . 1955. *Studien zur mittelbabylonischen Grammatik. Studia Orientalia* 20. Helsinki: Societas Orientalis Fennica.
- . 1959. "Die semitische Zischläute (*t*), *š*, *s* und *s* und ihre Vertretung im Akkadischen." *Orientalia* 28: 321–35.
- Berkooz, Moshé. 1937. *The Nuzi Dialect of Akkadian: Orthography and Phonology* (Language Dissertations 23). Philadelphia: University of Pennsylvania.
- Bomhard, Allan R. 1988. "The Reconstruction of the Proto-Semitic Consonant System." In *Fucus: A Semitic/Afrasian Gathering in Remembrance of Albert Ehrman*, ed. Yoël L. Arbeitman (Current Issues in Linguistic Theory 58), pp. 113–40. Amsterdam: Benjamins.
- Buccellati, Giorgio. 1966. *The Amorites of the the Ur III Period* (Ricerche 1). Naples: Istituto Orientale di Napoli.
- . 1971. Review of Jucquois 1966. *Oriens Antiquus* 10: 79–83.
- . 1979. "Comparative Graphemic Analysis of Old Babylonian and Western Akkadian." *Ugarit-Forschungen* 11: 89–100.
- . 1982. "Studies in Ebla Graphemics, 1." *Studi Eblaiti* 5: 39–74.
- . 1984. "Introduction." In *Graphemic Categorization, 2: The Middle Assyrian Laws*, by Claudio Saporetti, pp. 1–20. Malibu, Calif.: Undena.
- . 1990a. "The Ebla Electronic Corpus: Graphemic Analysis." *Annales Archéologiques Arabes Syriennes* 40: 8–26.
- . 1990b. "From Khana to Laqê: The End of Syro-Mesopotamia." In *De la Babylonie à la Syrie, en passant par Mari (Mélanges Kupper)*, ed. Ö. Tunca, pp. 229–53. Liège: Université de Liège.
- . 1990c. "'River Bank,' 'High Country' and 'Pasture Land': The Growth of Nomadism on the Middle Euphrates and the Khabur." In *Tall al-Hamidiyah* 2, ed. S. Eichler, M. Wäfler, and D. Warbuton, pp. 87–117. Göttingen: Vandenhoeck & Ruprecht.
- . 1992. "Ebla and the Amorites." *Eblaitica* 3: 85–106.
- . forthcoming. *A Structural Grammar of Babylonian*. Wiesbaden: Harrassowitz.
- Cantineau, Jean. 1951–52. "Le consonantisme du sémitique." *Semitica* 4: 79–94.
- Daniels, Peter T. 1993. "Is a Structural Graphemics Possible?" *Eighteenth LACUS Forum* 528–37.
- Degen, Rainer. 1969. *Altaramäische Grammatik der Inschriften des 10.–8. Jh. v. Chr.* (Abhandlungen für die Kunde des Morgenlandes 38/3). Wiesbaden: Steiner.

- Deller, Karlheinz. 1959. "Lautlehre des Neuassyrischen." Ph.D. dissertation, University of Vienna.
- . 1962a. "Studien zur neuassyrischen Orthographie." *Orientalia* 31: 186–96.
- . 1962b. "Zweisilbige Lautwerte des Typs KVKV im Neuassyrischen." *Orientalia* 31: 7–26.
- De Meyer, Leon. 1962. *L'Accadien des contrats de Suse* (Suppléments Iranica Antiqua 1). Leiden: Brill.
- Diakonoff, Igor M. 1981. "Towards the Pronunciation of a Dead Language." *Assyriological Miscellanies* (Copenhagen) 1: 7–12.
- . 1985. "The Importance of Ebla for History and Linguistics." *Eblaitica* 2 (1990): 3–29.
- . 1991–92. "Proto-Afrasian and Old Akkadian: A Study in Historical Phonetics," with Contributions by Olga Stolbova and Alexander Militarev. *Journal of Afroasiatic Languages* 4: 1–133.
- Edzard, Dietz Otto. 1986. Review of Greenstein 1984. *Journal of the American Oriental Society* 106: 359–62.
- Faber, Alice. 1985. "Akkadian Evidence for Proto-Semitic Affricates." *Journal of Cuneiform Studies* 37: 101–07.
- Finet, André. 1956. *L'Accadien des lettres de Mari* (Classe des lettres et des sciences morales et politiques, Mémoires 81). Brussels: Académie royale de Belgique.
- Gaebelein, Paul W. 1976. "Graphemic Analysis of Old Babylonian Letters from Mari." Ph. D. dissertation, University of California, Los Angeles.
- Garbini, Giovanni. 1960. *Il semitico di Nord-Ovest* (Quaderni della Sezione Linguistica degli Annali 1). Naples: Istituto Universitario Orientale di Napoli.
- . 1965. "Configurazione dell'unità linguistica semitica." In *Le "Protolingue": Atti del IV convegno internazionale di linguisti tenuto a Milano nei giorni 2–6 settembre 1963, presso il Sodalizio Glottologico Milanese* (Atti del Sodalizio glottologico milanese, Supplemento 16). Milan and Brescia: Paideia.
- . 1972. *Le lingue semitiche* (Studi di storia linguistica, Ricerche 9). Naples: Istituto Orientale di Napoli.
- Geers, Frederic W. 1965. "The Treatment of Emphatics in Akkadian." *Journal of Near Eastern Studies* 4: 65–67.
- Gelb, I. J. 1952. *Old Akkadian Writing and Grammar* (Materials for the Assyrian Dictionary 2). Chicago: University of Chicago Press (See Gelb 1961).

- . 1957. *Glossary of Old Akkadian* (Materials for the Assyrian Dictionary 3). Chicago: University of Chicago Press.
- . 1958. "La lingua degli Amoriti." *Atti della Accademia nazionale dei Lincei: Rendiconti della Classe di scienze morali, storiche e filologiche*, ser. 8, vol. 13, no. 3–4: 143–64.
- . 1961. *Old Akkadian Writing and Grammar* (Materials for the Assyrian Dictionary 2), 2nd ed., pp. 24–43, 119–27. Chicago: University of Chicago Press.
- . 1970a. "Comments on the Akkadian Syllabary." *Orientalia* 39: 516–46.
- . 1970b. "A Note on Morphographemics." In *Mélanges Marcel Cohen*, ed. D. Cohen (Janua Linguarum Series Major 27), pp. 73–77. The Hague: Mouton.
- . 1980. *Computer-aided Analysis of Amorite*, with the assistance of J. Bartels, S.-M. Vance, and R. M. Whiting (Assyriological Studies 21). Chicago: Oriental Institute.
- Gelb, I. J., and Burkhardt Kienast. 1990. *Die altakkadischen Königsinschriften des dritten Jahrtausends v. Chr.* (Freiburger Altorientalische Studien 7). Stuttgart: Steiner.
- Giakumakis, George, Jr. 1970. *The Akkadian of Alalah* (Janua Linguarum Series Practica 59). The Hague: Mouton.
- Goetze, Albrecht. 1937. "The Sibilant in Old Babylonian *naṣārūm*." *Orientalia* 6: 12–18.
- . 1946. "Sequence of Two Short Syllables in Akkadian." *Orientalia* 15: 233–38.
- . 1947. "Short or Long *ā*?" *Orientalia* 16: 239–50.
- . 1958. "The Sibilants of Old Babylonian." *Revue d'Assyriologie* 52: 137–49.
- Greenfield, Jonas C. 1969. "Amurrite, Ugaritic and Canaanite." In *Proceedings of the International Conference on Semitic Studies Held in Jerusalem 19–23 July 1965*, pp. 92–101. Jerusalem: Israel Academy of Sciences and Humanities.
- Greenstein, Edward L. 1977. "Phonological Studies in Akkadian." Ph.D. dissertation, Columbia University.
- . 1980. "The Assimilation of Dentals and Sibilants with Pronominal *š* in Akkadian." *Journal of the Ancient Near Eastern Society* 12: 51–64.
- . 1984. "The Phonology of Akkadian Syllable Structure." *Afroasiatic Linguistics* 9/1: 1–71.

- Haupt, Paul. 1884–85. "Assyrian Phonology, with Special Reference to Hebrew." *Hebraica* 1: 175–81.
- . 1887. "The Assyrian E-Vowel: A Contribution to the Comparative Phonology of the Assyro-Babylonian Language." *American Journal of Philology* 8: 261–91.
- . 1887. "Über den Halbvocal *w* im Assyrischen." *Zeitschrift für Assyriologie* 2: 259–86.
- . 1890. "Über die beiden Halbvocale *w* und *y*." *Beiträge zur Assyriologie* 1: 293–300.
- Hecker, Karl. 1968. *Grammatik der Kültepe-Texte* (Analecta Orientalia 44). Rome: Pontificium Institutum Biblicum.
- Huehnergard, John. 1988. *The Akkadian of Ugarit* (Harvard Semitic Studies 34), pp. 23–122. Atlanta: Scholars Press.
- Huffmon, Herbert B. 1965. *Amorite Personal Names in the Mari Texts: A Structural and Lexical Study*. Baltimore: The Johns Hopkins University Press.
- Hyatt, James Philip. 1941. *The Treatment of Final Vowels in Early Neo-Babylonian* (Yale Oriental Series, Researches 23). New Haven: Yale University Press.
- Izre'el, Shlomo. 1987. "The Complementary Distribution of the Vowels *e* and *i* in the Peripheral Akkadian Dialect of Amurru: A Further Step towards Our Understanding of the Development of the Amarna Jargon." In *Proceedings of the Fourth International Hamito-Semitic Congress (Marburg, 20–22 September 1983)*, ed. H. Jungraithmayr and W. W. Müller (Current Issues in Linguistic Theory 44), pp. 525–41. Amsterdam: Benjamins.
- . 1991. *Amurru Akkadian: A Linguistic Study*, with an appendix on the history of Amurru by Itamar Singer, 2 vols. (Harvard Semitic Studies 40–41). Atlanta: Scholars Press [esp. 1:26–94].
- Jäger, M. 1890a. "Der Halbvokal *y* im Assyrischen." *Beiträge zur Assyriologie* 1: 443–91.
- . 1890b. "Das babylonische Hauchlautzeichen." *Beiträge zur Assyriologie* 1: 589–92.
- Jucquouis, Guy. 1966. *Phonétique comparée des dialects moyen-babyloniennes du nord et de l'ouest* (Bibliothèque du Muséon 53). Louvain: Institut Orientaliste.
- Kinnier Wilson, J. V. 1968. "'Desonance' in Accadian." *Journal of Semitic Studies* 13: 93–103.

- Knudsen, Ebbe E. 1961. "Cases of Free Variants in the Akkadian *q* phoneme." *Journal of Cuneiform Studies* 15: 84–90.
- . 1969. "Spirantization of Velars in Akkadian." In *Lišān mithurti: Festschrift von Soden*, ed. W. Röllig (Alter Orient und Altes Testament 1), pp. 147–56. Neukirchen-Vluyn: Neukirchener Verlag.
- . 1980. "Stress in Akkadian." *Journal of Cuneiform Studies* 32: 3–16.
- . 1982a. "An Analysis of Amorite." *Journal of Cuneiform Studies* 34: 1–18 [review of Gelb 1980].
- . 1982b. "The Mari Akkadian Shift *ia* > *ê* and the Treatment of 𐎶𐎵 Formations in Biblical Hebrew." *Journal of Near Eastern Studies* 41: 35–43.
- . 1986. Review of Greenstein 1984. *Bibliotheca Orientalis* 43: 723–32.
- . 1991. "Amorite Grammar: A Comparative Statement." In *Semitic Studies in Honor of Wolf Leslau on the Occasion of his Eighty-fifth Birthday, November 14th, 1991*, ed. Alan S. Kaye, pp. 866–85. Wiesbaden: Harrassowitz.
- Kobayashi, Yoshitaka. 1975. "Graphemic Analysis of Old Babylonian Letters from South Babylonia." Ph.D. dissertation, University of California, Los Angeles.
- Labat, René. 1946. "Le système phonétique de l'akkadien: Observations préliminaires à une étude phonologique." *Comptes Rendues du Groupe Linguistique des Etudes Chamito-Sémitiques* 4: 9–12.
- Lieberman, Stephen J. 1977. *The Sumerian Loanwords in Old-Babylonian Akkadian*, vol. 1: *Prolegomena and Evidence*. Missoula, Mont.: Scholars Press.
- . 1979. "The Phoneme /o/ in Sumerian." In *Studies in Honor of Tom B. Jones*, ed. M. A. Powell and R. H. Sack, pp. 21–28. Neukirchen-Vluyn: Neukirchener Verlag.
- Mayer, Walter. 1971. *Untersuchungen zur Grammatik des Mittelassyrischen* (Alter Orient und Altes Testament Sonderreihe 2). Neukirchen: Neukirchener Verlag.
- Mendenhall, George E. 1985. *The Syllabic Inscriptions from Byblos*. Beirut: American University of Beirut.
- Murtonen, A. 1966. "The Semitic Sibilants." *Journal of Semitic Studies* 11: 135–50 [esp. pp. 149f.].
- Platt, James H. 1993. "Eblaite Scribal Schools: Graphemics and Orthography." Ph.D. dissertation, University of California, Los Angeles.

- Poebel, Arno. 1939. *Studies in Akkadian Grammar* (Assyriological Studies 9). Chicago: University of Chicago Press.
- Polomé, Edgar C. 1981. "Can Graphemic Change Cause Phonemic Change?" In *Bono Homini Donum: Essays in Historical Linguistics in Memory of J. Alexander Kerns*, ed. Yoël L. Arbeitman and Allan R. Bomhard (Current Issues in Linguistic Theory 16), vol. 2, pp. 881–88. Amsterdam: Benjamins.
- Prosecky, Jiri. 1986. "L'analyse graphémique de la version paléobabylonienne du mythe d'Atramhasis: Contribution à la méthode de l'analyse graphémique des textes cunéiformes akkadiens." *Archiv Orientalní* 54: 61–76.
- Rabin, Chaim. 1969. "The Structure of the Semitic System of Case Endings." In *Proceedings of the International Conference on Semitic Studies Held in Jerusalem 19–23 July 1965*, pp. 190–204. Jerusalem: Israel Academy of Sciences and Humanities.
- Ravn, Otto Emil. 1939. "The Substitution of -ss- for -šš- in Babylonian." *Acta Orientalia* 17: 318–28.
- Reiner, Erica. 1964. "The Phonological Interpretation of a Subsystem in the Akkadian Syllabary." In *Studies Presented to A. Leo Oppenheim*, pp. 167–80. Chicago: Oriental Institute.
- . 1966. *A Linguistic Analysis of Akkadian* (Janua Linguarum Series Practica 21). The Hague: Mouton.
- . 1970. "Akkadian." In *Current Trends in Linguistics*, ed. Thomas A. Sebeok, vol. 6, *Linguistics in South West Asia and North Africa*, pp. 274–303. The Hague: Mouton.
- . 1973a. "How We Read Cuneiform Texts." *Journal of Cuneiform Studies* 25: 3–58.
- . 1973b. "New Cases of Morphophonemic Spellings." *Orientalia* 42: 35–38.
- . 1985. "The Uses of Astrology." *Journal of the American Oriental Society* 105: 589–95.
- Riemschneider, Kaspar K. 1976. "Compound Graphemic Units in Standard Babylonian Writing." *Journal of Cuneiform Studies* 28: 65–71.
- Rimalt, E. S. 1933–34. "Zur Lautlehre des Neubabylonischen." *Archiv für Orientforschung* 9: 124–26.
- Salonen, Erkki. 1962. *Untersuchungen zur Schrift und Sprache des altbabylonischen von Susa* (Studia Orientalia 27/1). Helsinki: Societas Orientalis Fennica.

- Sarauw, Chr. 1939. *Über Akzent und Silbenbildung in den älteren semitischen Sprachen*. Copenhagen.
- Segert, Stanislav. 1975. *Altaramäische Grammatik mit Bibliographie, Chrestomathie und Glossar*. Leipzig: VEB Verlag Enzyklopädie.
- Shehadeh, Lamia R. 1987. "Some Observations on the Sibilants in the Second Millennium B.C." In *Working with No Data: Semitic and Egyptian Studies Presented to Thomas O. Lambdin*, ed. David M. Golomb, pp. 229–46. Winona Lake, Ind.: Eisenbrauns.
- Steiner, Richard. 1977. *The Case for Fricative-Laterals in Proto-Semitic* (American Oriental Series 59). New Haven: American Oriental Society [esp. pp. 144–48, 159].
- . 1982. *Affricated Šade in the Semitic Languages* (American Academy for Jewish Research Monograph Series 3). New York: American Academy for Jewish Research [esp. pp. 70–74].
- von Soden, Wolfram. 1948a. *Das akkadische Syllabar* (Analecta Orientalia 27). Rome: Pontificium Institutum Biblicum.
- . 1948b. "Vokalfärbungen im Akkadischen." *Journal of Cuneiform Studies* 2: 291–303.
- . 1952. *Grundriss der akkadischen Grammatik* (Analecta Orientalia 33). Rome: Pontificium Institutum Biblicum.
- . 1965–81. *Akkadisches Handwörterbuch*. Wiesbaden: Harrassowitz.
- . 1968. "Die Spirantisierung von Verschlusslauten im Akkadischen: Ein Zwischenbericht." *Journal of Near Eastern Studies* 27: 214–20.
- . 1969. *Grundriss der akkadischen Grammatik samt Ergänzungsheft* (Analecta Orientalia 33/47). Rome: Pontificium Institutum Biblicum.
- von Soden, Wolfram, and Wolfgang Röllig. 1967. *Das akkadische Syllabar*, 2nd ed. (Analecta Orientalia 42). Rome: Pontificium Institutum Biblicum.
- Westenholz, Aage. 1978. "Some Notes on the Orthography and Grammar of the Recently Published Texts from Mari." *Bibliotheca Orientalis* 35: 160–69 [esp. pp. 161–64].
- Wilhelm, Gernot. 1970. *Untersuchungen zum Hurro-Akkadischen von Nuzi* (Alter Orient und Altes Testament 9). Neukirchen-Vluyn: Neukirchener Verlag.
- . 1983. "Reconstructing the Phonology of Dead Languages." In *Writing in Focus*, ed. F. Coulmas and K. Ehlich (Trends in Linguistics, Studies and Monographs 24), pp. 157–66. Berlin: Mouton de Gruyter.

Zadok, Ran. 1993. "On the Amorite Material from Mesopotamia." In *The Tablet and the Scroll: Near Eastern Studies in Honor of William W. Hallo*, ed. M. E. Cohen, D. C. Snell, and D. B. Weisberg, pp. 315–33. Bethesda, Md.: CDL Press.

Zimmern, Heinrich. 1890. "Zur assyrischen und vergleichende semitischen Lautlehre." *Zeitschrift für Assyriologie* 5: 367–98.

Eblaite Phonology

Cyrus H. Gordon

New York University

Eblaite designates the language of the largest archives known from anywhere in the world during the Early Bronze Age (3000–2000 B.C.E.). The inscriptions of the Ebla archives already unearthed number about 15,000, many on large, well-preserved clay tablets. They have been excavated by an Italian expedition at the ancient site of Ebla, a little over thirty miles south of Aleppo, Syria. The archives were written during a period of about fifty years during the 23rd century.

Eblaite is a Semitic language embodying East and West Semitic features. Texts in the same language have been found at Mesopotamian sites such as Mari, Tell Abū-Šalābīḥ, and Kish. Eblaite is essentially a written *lingua franca* used by scribes, merchants, and diplomats. It was not limited to Ebla, nor was it the spoken language of Ebla.

Scribes, and some of their employers, could probably converse in it, but it is most unlikely that the scribes, merchants, diplomats, or anyone else spoke Eblaite at home.

The Eblaite texts are written in the cuneiform system of Mesopotamia, developed by the Sumerians and adopted by the Semitic Akkadians. The system has ideograms in addition to syllabic signs. It also employs determinatives to fix the semantic category of a word. So many Sumerograms are used that it is often possible for an Assyriologist to sense the meaning of a passage, or even of an entire tablet, without knowing how to pronounce it in Eblaite. The ends of words are sometimes added syllabically to the Sumerograms so that we can deduce the case endings of nouns and the modal suffixes of verbs, as well as the suffixed personal pronouns and conjugational suffixes. Particles (conjunctions, prepositions, and the like) and names are usually spelled out phonetically. There are also extensive bilingual school texts giving us the Eblaite translations of Sumerograms.

AUTHOR'S NOTE: Abbreviations: DN, divine name; GN: geographical name; PN, personal name.

However, much of that vocabulary is highly specialized and not as applicable as we might wish to the other tablets found by the excavators.

In the Ebla archives, the Sumerograms do not follow Sumerian word order, but instead are placed in accordance with the rules of Eblaite syntax. We can thus learn much about Eblaite phrase and sentence structure even when we cannot pronounce the Eblaite words for which the Sumerograms stand.

The reconstruction of Eblaite does not involve the decipherment of a script, nor essentially the interpretation of the texts. It is rather the extracting, analyzing, classifying, and assembling of seemingly endless linguistic details with the aim of enabling us to compose a grammar and glossary of the Eblaite language. For the on-going process, see Archi (1987), Diakonoff (1990), and Gordon (1987, 1990, 1992).

Eblaite is written in the Mesopotamian cuneiform system which was not designed for, and hence is not well suited for, recording Semitic languages. In spite of nearly a century and a half of linguistic scholarship devoted to Akkadian, many details of the phonology remain controversial. The script is characterized by the two opposing (and confusing) principles of polyphony and homophony. Polyphony means that a sign may have several (often many) different phonetic values; e.g., there is a common sign that is to be pronounced *ur*, or *taš*, or *lik* (along with several other values) depending on context. Homophony means that the same sound may be represented by several (often many) entirely different signs.

When whole, phonetically spelled Eblaite texts will have come to light, no seasoned Semitist should experience insuperable trouble in reading, translating, and describing them linguistically in detail. This will be the case no matter whether the script is syllabic or alphabetic. Our main present obstacle is simply that there is no sizable Eblaite prose or poetic literature spelled out phonetically without a plethora of Sumerograms and proper names. Names are often unconnected with the language of the people who bear them. My name (Cyrus) is Persian and my wife's name (Constance) is Latin (Constantia). Neither of us is Iranian or Italic and our language is English.

The Eblaite system of writing is inherited from Sumerian proto-writing of Early Dynastic II. Semitic phonology includes a threefold repertoire of dental and palatal stops and sibilants, namely *surd* (voiceless), *sonant* (voiced), and *emphatic*. Thus we find *t/d/ṭ*, *k/g/q*, and *s/z/š*. Sumerian lacks the emphatic; so that *DA* covers *da* and *ṭa*, *GA* covers *ga* and *qa*, and *ZA* covers *za* and *ša*. However, in Eblaite all three grades are lumped together so that *DA* covers *ta/da/ṭa*, *GU* covers *ku/gu/qu*, and *SA* covers *sa/za/ša*. Neither the doubling of consonants nor the length of vowels is normally indicated.

It follows that normalization and etymology cannot be ascertained mechanically. The process requires an extensive and systematic knowledge of the Semitic languages.

In the bilinguals, ŠU-TUR (literally ‘little hand’) designates ‘finger’, and its Eblaite form corresponds to the Common Semitic word that appears as ‘ešba’ ‘finger, toe’ in Hebrew. The Eblaite is written two ways: *i-sa-ba-um* TUR and *iš-ba-um* TUR. Note the open-syllable spelling with *sa* for vowelless [š] in the first form, and also note the *š* for *s* in the second form.

Eblaite is often written in open-syllable orthography (like Linear A and B, or Japanese). A word beginning [ʾašmi-] is spelled either *a-za-mi-* or *a-zi-me-*; note that the vowelless [š] is written either *za*, reflecting the vowel of the preceding syllable, or *zi*, reflecting the *i* of the following syllable. This open-syllable orthography often gives the illusion that there are no closed syllables in Eblaite, which is not at all the case.¹

The mixed character of Eblaite is reflected in the words designating the large numbers: *mi-at* ‘100’ is Common Semitic, *li-im* ‘1,000’ is East Semitic, *rī-bab* ‘10,000’ is West Semitic, and *ma-i-at* ‘100,000’ goes its own way in a new direction.

It is interesting to note the Eblaite forms of Sumerian names: ^dEN-LIL > *I-li-lu*, ^dNIN-KAR-DU > *Ni-ka-ra-du*, ^dSUMUQAN > *Sa-ma-gan*, ^dAŠNAN > *A-sa-ma-an*.

The bilingual equation BAḤAR ‘potter’ = *wa-šī-lu-um* (cognate with Hebrew *yôšēr* ‘potter’) illustrates that (1) unlike in Northwest Semitic, initial [*w-] remains in Eblaite (as in Akkadian and Arabic) without shifting to [y-]; and (2) [r] may change to [l]. The change of [r] to [l] (which is quite common) is not reversible, for [l] never shifts to [r]. The falling together of [r] and [l] (in the [r > l] shift) reflects a common phenomenon. In Linear A and B, and in hieroglyphic Egyptian, [r] and [l] fall together, at least in the orthography. In Chinese and Japanese, they definitely fall together in the spoken language as well.

The GN *Ar-ga*^{ki} [ʾarqa] has lost the final *-t* of the singular feminine suffix *-at*. In the Amarna letters this GN is called *Arqat*. Cf. the GN *É-ma*^{ki} = *Ḥāmāt*

1. Gimbutas (1991) has shown that the Linear A graphs are derived from the Old European graphs of about 5300–4300 B.C.E. See chap. 8, “The Sacred Script” (esp. p. 320, fig. 8-22). She attributes the disappearance of the Old European script toward the end of the fifth millennium to the Indo-Europeanization of Central Europe at that time. The people who had fostered the Old European script were forced into the Aegean area, carrying that script to Crete and Cyprus, where it survived for millennia. Her discovery will have a major impact on the reconstruction of ancient history, in several different ways.

(modern Hama). In the *status absolutus* the final [-t] of the singular feminine suffix is generally dropped in Hebrew, and regularly in spoken Arabic. Final [-t] begins to be dropped in Egyptian already in the third millennium.

The GN *Māš-a*^{ki} is to be compared with the GN *Maššā'* (Prov. 30:1, 31:1). The -a accordingly does not reflect the feminine suffix (which would be written with a final -h in Hebrew), but final -'.

Etymology indicates that [l] frequently loses its normal consonantal character in Eblaite and is not represented in the script. E.g., *a-bi-nu-u(m) i-a-ba-nu* SIG₄-GAR [ābinū yabānū libitta] 'brickmakers will make the brick'. The root of the first two words is **lbn* although no [l] appears in the orthography. We must reckon with vocalic [l] (and also vocalic [r]) in the Semitic languages. In Akkadian, the second radical of all quadriconsonantal verbs is either [l] or [r]. This may explain the Arabic IIIrd conjugation (*qātala*; cf. Heb. *pōlēl* with a long vowel representing the absorption of vocalic [l] or [r]).

Initial [w-] does not undergo the Northwest Semitic shift to [y-]. Eblaite retains [w-] like Akkadian, Arabic, and Ethiopic. In addition to [w-] (discussed above), note (the dual) *wa-ti-a* [wādi-a] 'the two wadis'.

The script reflects four vowels: *a*, *i*, *u/o*, and *e*. However, *e* may well be non-phonemic but only positionally conditioned by contact with *ḥ* or '.

The diphthongs [ay], [aw], and [ue] can all be reduced to [ā]. In the case of [ay > ā] and [aw > ā], the shift can be explained as "falling" diphthongs (with the accent on the [a]) as distinct from "rising" diphthongs (with the accent on the [y/w]) whereby **bayt* yields *bīt* 'house' and **mawt* yields *mūt* 'death' in Akkadian. While we find many Eblaite examples of [ay > ā] and several of [aw > ā], there is only one of [ue > ā]: the Akkadian *Suen* (which comes into Babylonian and Assyrian as *Sin*) is reduced to *San* 'Moon' in *San-Ugāru* 'Moon of the Field' (an epithet of the Moon) in an incantation where the reference to the Moon is fixed by duplicate passages which have *ṛṛ* 'month, moon' instead. *San* also appears instead of *Sin* (the Moon god) in the Hebrew form of the Mesopotamian names *San-ḥērīb* (Sennacherib) and *San-ballat*. In the Palestinian GN 'Beth-shean' the second element (Shan) designates the Moon god. Examples of [ay > a] in Eblaite: *ba-du* [bāt-] 'house'; *a-na* and *a-na-a* ['anā] 'eyes' (dual); *ba-nu* [bān-] 'tamarisk' (vs. Akkadian *bīn-* with rising diphthong); *ma-sa-lu-u(m)* [masalu] 'justice, uprightness' from **yšr* (note Hebrew *mēšārīm*) and from the Št stem of the same root *uš-da-ši-ir* [uštašir] 'he prepared, released (lit. caused to be right)'. There are fewer examples of [aw > ā]: [*'aw > 'ā] 'or'; and the PN *Mu-ša-ra-du* (Š participle of **wrd* 'to bring down gods with gifts' = 'to propitiate them

successfully'; cf. Arabic and especially Ugaritic).

Cryptic writings were meant to be read by the initiated. We know from a bilingual tablet that ^dMUL = *Kab-kab* 'Star'; however, in an incantation it is written *ga : ga : ba : bù* [kabkabu] (ending in nominative -u). A similar cryptograph is *ga : ga : li : la*, which calls to mind *galgal* 'wheel'; if correct, note that *li* stands for vowelless [l].

The name *Da-gu-nu* corresponds to Heb. *Dāgôn* (< *Dagān*) with ā > ō, which is frequent in Canaanite. Cf. also Minoan *Da-gu-na*.

[-dk-] is assimilated to [-kk-] in *a-za-me-ga* ['ašmikka < *'ašmid-ka] 'I bind thee'.

Sum-ar-rum^{ki} 'Sumer' comes into Hebrew as *Šin'ār* 'Babylonia'. Note that AR = 'ar and begins a new syllable. Another change in syllabification is inherent in Eblaite and Old Akkadian *en-ma* (> standard Akkadian *um-ma*) and Hebrew [נַם]—all meaning 'so says' (followed by direct discourse). The initial *n*- is properly vocalic [n̥], which is written either *en-* or *um-* according to the rules of the Mesopotamian syllabary, and 𐎠 (*nē*) according to the orthographic rules of Masoretic pointing.

Vocalic nasals and liquids include [r]. The tree called the ^{giš}ŠU-ME is bilinearly rendered in Eblaite as either *šī-rí-mi-nu* or *ša-mi-nu*. Expressed alphabetically, the choice is between *šrmn* and *šmn*. The [r] is vocalic and not represented in the latter.

Philippi's law ([i > a] in an originally closed accented syllable) is operative in Northwest Semitic (Canaanite, Aramaic, Ugaritic). It does not take place in Akkadian, Arabic, etc. Since Eblaite is a border language between East and Northwest Semitic, it is not surprising that it occurs sporadically. Thus in Eblaite both *libittu* and *libattu* (spelled *li-bi-tum* and *li-ba-tum*) 'brick' occur.

The following independent personal pronouns are attested and bring out the difference between the orthography and probable phonetic pronunciation. In the case of *a-na/an-na* 'I', ['anna] is preferred (vs. the pronoun in the other Semitic languages) because in normal Mesopotamian orthography, doubling a consonant means phonetic doubling, while writing a single consonant can stand for either a single or a double consonant in the pronunciation.

sg. 1 comm.nom.	<i>an-na, a-na</i> ['anna] 'I'
sg. 2 masc.nom.	<i>an-da</i> ['anta] 'thou'
acc.	<i>gu-wa-ti</i> [kuwāti] 'thee'
dat.	<i>gu-wa-si</i> [kuwāši(m)] 'to thee'

sg. 3 masc.nom.	<i>su-wa</i> [šuwa] ‘he’
acc.	<i>su-wa-ti</i> [šuwati] ‘him’
dat.	<i>su-wa-si</i> [šuwāši(m)] ‘to him’
sg. 3 fem. nom.	<i>si-a</i> [šiya] ‘she’
du. 1 comm.dat.	<i>ne-si-in</i> [nešin] ‘to both of us’
pl. 2 masc.nom.	<i>an-da-nu</i> [ʾantanu] ‘ye, you’
pl. 3 masc.nom.	<i>su-nu</i> [šunū] ‘they’

The following suffixed personal pronouns are attested:

sg. 1 comm.gen.	-i [-ī], or (postvocalic) -a [-(y)a] ‘my’
acc.	-ni [-ni] ‘me’
sg. 2 masc.gen./acc.	-ga [-ka] ‘thy, thee’
dat.	-kum [-kum] ‘to thee’
sg. fem. gen./acc.	-gi [-ki] ‘thy, thee’
sg. 3 masc.gen./acc.	-sù, -su [-šu] ‘his, him’
dat.	-su-um [-šum] ‘to him’
sg. fem. gen.	-sa [-ša] ‘her’
pl. 1 comm.gen.	-na [-nā], -nu [-nū] ‘our’
pl. 2 masc.acc.	-gu-nu [-kunū] ‘you’
pl. 3 masc.gen./acc.	-su-nu [-šunū] ‘their, them’
pl.3 fem. gen.	-si-na [-šina] ‘their’
acc.	-si-na-at [-šināt] ‘them’

The lone occurrence of an unusual pronominal suffix calls for special notice. The meaning of *ši-ne-mu* [šinn-êmo] ‘his teeth’ is fixed by context. That all three vowels are to be pronounced is indicated by the difference among [i, e, u], which rules out zero-grade vowels disguised by open-syllable orthography. The ending *-êmo*, known from biblical Hebrew poetry, is a general possessive suffix that can be applied regardless of person. Here it means ‘his teeth’ but in another context it could mean ‘their teeth’ and so forth. An exact (though admittedly clumsy) translation is ‘thereof’.

The six case endings are: nominative *-u(m)*, genitive *-i(m)*, accusative *-a(m)*, dative-locative *-iš*, locative-adverbial *-u(m)*, absolute *-a/∅*. The three commonest cases (nom., gen., acc.) are often confused. There are some indeclinables in *-a*, especially among the PNs such as *Ra-ba*, *Ba-ga-ma*, *Tab-rí-sá* and the DN ^d*Ba-ra-ma*. While a bilingual renders the name of the Sumerian god EN-KI as *É-um* [Ḥayyum] ‘The Living One’ in Eblaite, the other Eblaite texts regularly render the name *É-a* [Ḥayya] ending in *-a*. Other DNs in *-a* are: *É-da* [Hadda] and *Qu-ra* [Qūra].

The absolute in [-Ø] (zero) is common in some proper names; e.g., the DN^s ^d*Ga-mi-iš*, ^d*Ra-sa-ap*; the GNs *A-da-bi-ik*^{ki}, *A-da-ti-ik*^{ki}, and month names such as ITI *za-É-na-at*.

The construct state is not always expressed in writing; e.g., it is not expressed in *ḥa-za-nu* GN ‘the mayor of X’, while it is in *ma-lik* GN ‘the advisor of Y’.

The suffix marking the dual is *casus rectus* -*ā(n)*, *obliquus* -*ay(n)* (which can shift to -*ā(n)* in accordance with the reduction of the diphthong [ay > a]); e.g., *tal-da-an* [daltān] ‘double doors’ and *su-lu-la-a* ‘the two horns’. The following citation has the dual suffix in the word designating a pair of deified rivers and a dual noun in apposition with them: 2 ^d*Ba-li-ḥa wa-ti-a* [Baliḥa wādia] ‘the two divine Baliḥ rivers, the two streams’.

There is one pair of words that raises fundamental questions. The nouns *ḥrd* ‘child’ and *ms* ‘son, child’ have long been known only from Egyptian. Then both turned up in Ugaritic of the Late Bronze Age, and now in Eblaite of the Early Bronze Age. Ebla had connections with Egypt; alabaster vessels with the names of Chefred (4th Dynasty) and Pepi I (6th Dynasty) have been found at Ebla in the archeological stratum that yielded the archives. The meaning of Eblaite *ḥar-da-du* [ḥardātu] in the sense of ‘young women’ is fixed by context; the same form with the same sense occurs in Old Kingdom Egypt. The situation with Eblaite *maš* (= Egyptian *ms*) is more complex and tantalizing. It is common in all periods of Egyptian from start to finish. But it also may occur in Sumerian (MÁŠ) with the meaning of ‘kid, young goat’. Words for young animals are often applied to children.

The adjective has long been known to be inflected like the noun except for the feminine plural, *status absolutus* in Syro-Aramaic. Now Ebla shares that exception with Syro-Aramaic. In Hebrew *mēlākîm/n* ‘kings’ has its feminine counterpart in *mēlākôt* ‘queens’. But in Aramaic the phonetic equivalent of Hebrew feminine plural construct *malkôt* מַלְכֹּת ‘queens (of)’, namely *malkât* מַלְכַּת ‘queens (of)’, can only serve as the construct. The Aramaic absolute (corresponding in meaning to Hebrew *mēlākîm/n* ‘kings’) is *malkān* מַלְכָּן ‘queens’. Thus Eblaite *du-na-an* [dunnān] ‘mighty (females)’ has the suffix -*ān* for the adjective plural feminine absolute.

There are two principal tenses: (1) the so-called imperfect with prefixes and some suffixes, and (2) the so-called perfect with suffixes but no prefixes. Verbs are usually classified as strong (with a root of three stable consonants) or weak (with a semivowel: *w* or *y*, functioning as one of the root

consonants; or with only two consonants in the root with the second one repeated). A root can be treated within the matrix of several conjugations, all familiar from the other Semitic languages; e.g., G, Gt, D, Š, ŠD, ŠDt. Like Aramaic, Eblaite has no N conjugation. The imperfect has modal suffixes; thus *-u* is the sign of the indicative. In the perfect, the singular 3rd masculine ends in *-a*. In Akkadian the perfect is for stative or intransitive verbs; in West Semitic the perfect is used for transitive as well as intransitive verbs; here, Eblaite goes with West Semitic.

The imperfect has *a*, *i*, or *u* as the thematic vowel between the last two consonants of the root:

u-class: *Iq-bu-ul-(Ma-lik)* is a PN meaning '(The Divine King) has accepted (the propitiatory offerings of the parents and granted the child who bears the PN)'.

i-class: *Ig-ri-iš-(Li-im)* is a PN meaning '(God) has driven out (the forces of evil)'.

a-class: *Ir-kab-(Ar)* is a PN meaning '(The deity Ar) rides'.

The prefix vowel gives the impression that there is an isogloss with East Semitic, in which we find Akkadian *iprus*, *iddin*, *išbat*, vs. West Semitic, where we find Arabic *yaqtul*, *yaʿlis*, *yasmaʿ*. Little weight can be attributed to the loss of the *y*- in Akkadian because of the vocalization of the Hebrew in the LXX (Septuagint Greek) tradition and some living traditions like the Arabic. Note LXX ΙΣΑΑΚ 'Isaac' (vs. Masoretic *Yiṣḥāq*) and Arabic *Iṣḥāq*; and vs. Masoretic *Yiśrāʾēl* 'Israel', note Arabic *Iśrāʾīl*. However, *ya-* is preserved in both the LXX and Arabic traditions: Masoretic *Yaʿaqōb*, LXX ΙΑΚΩΒ-, Arabic *Yaʿqūb*.

The attested morphs of the imperfect tense are: sg. 1st comm. *ʾa*-CCvC, sg. 2nd masc. *ta*-CCvC, sg. 3rd masc. *i*-CCvC, sg. 3rd fem. *ta*-CCvC, pl. 3rd masc. *i*-CCvC-*ū*.

The perfect (3rd masc. sg. = CaCvCa) is used in PNs such as *Ra-ga-ma-Il* 'God has spoken' (with **rgm* 'to speak' as in Ugaritic) or *Qá-ba-Lum* 'God has spoken' and *Qa-ba-Da-mu* '(The god) Damu has spoken' (where the verb is familiar from Akkadian *qabū* 'to speak'). East Semitic restricts the perfect to intransitive verbs, whereas Eblaite, like West Semitic, uses it for transitive as well as intransitive verbs. In addition to the above transitive perfects, note also *ba-na-a* 'he has built'.

The following is a D imperfect intensified by an infinitive absolute ending in adverbial *-u* (as in Ugaritic and Akkadian): *i-na-é-áš na-é-su* [*inaḥḥaš naḥāšu*] 'I shall verily perform magic'. Note that the D prefix does not go with Arabic and Akkadian *ʾu*-.

The infinitive absolute in *-u* appears in the *figura etymologica*: *ḥu-mu-zu ḥa-ma-zi, bû-ru₁₂ ba-ra-ru₁₁ (*brr)*.

The ŠD participle *muška^ainum* is of interest because it explains Akkadian *muškēnum* ‘helot, plebeian’, which survives into modern Hebrew *miskēn* and Arabic *miskīn* ‘poor’; and French *mesquin* ‘shabby, mean’.

The composite, interregional nature of Eblaite precluded the modicum of consistency and uniformity that more natural languages have developed through analogic leveling. We thus find the same root (**hlk*) with two treatments of the G infinitive in the bilinguals: *É-a-gu-um* [hākum] and *É-la-gum* [halākum] ‘to go’. There would be little merit in our striving to create a consistency that is not there.

The Early Bronze Age date of the Ebla archives provides an abundance of Semitic documents in Syria a millennium earlier than the Ugaritic tablets and half a millennium earlier than Minoan Linear A.

Bibliography

Sources

ARET: Archivi Reali di Ebla—Testi, Rome (the chief publication of the Ebla tablets; eight volumes have appeared).

Eblaitica: Eblaitica: Essays on the Ebla Archives and Eblaite Language (Publications of the Center for Ebla Research at New York University), ed. C. H. Gordon and Gary Rendsburg. Winona Lake, Ind.: Eisenbrauns. 1, 1987; 2, 1990; 3, 1992.

MEE: Materiali Epigrafici di Ebla, Naples (four volumes have appeared, including Giovanni Pettinato’s vol. 4 (two parts, 1982) on the bilinguals, which is very important).

SEB: Studi Eblaiti, University of Rome (the journal of the Ebla Expedition, headed by Paolo Matthiae; valuable articles have appeared in it).

UT: C. H. Gordon, *Ugaritic Textbook*. Rome: Pontifical Biblical Institute, 1967.

References

Archi, Alfonso. 1987. “Ebla and Eblaite.” *Eblaitica* 1: 7–17.

Diakonoff, I. M. 1990. “The Importance of Ebla for History and Linguistics.” *Eblaitica* 2: 3–29.

Gimbutas, Marija. 1991. *The Civilisation of the Goddess*. San Francisco: HarperCollins.

Gordon, C. H. 1987. "Eblaitica." *Eblaitica* 1: 19-28.

———. 1990. "Eblaite and Northwest Semitic." *Eblaitica* 2: 127-39.

———. 1992. "The Ebla Exorcisms." *Eblaitica* 3: 117-37.

Chapter 3
Ugaritic Phonology
Cyrus H. Gordon
New York University

Ugaritic preserves Semitic phonetic structure better than any known language except North and South Arabic. Our topic is complicated by the fact that Ugaritic is not a single dialect recorded in a single script. Although the Ugaritic texts date from a restricted period (ca. 1400–ca. 1185 B.C.E.) and most of them come from a restricted area (a small city-state along the coast of Syria north of Latakia), they are not in one homogeneous dialect. Prose and poetry often operate according to different rules, and there are within each of those two categories texts that deviate from their own norm in various ways, including phonetically.

Most of the Ugaritic tablets are written in an alphabet of thirty letters which, however, reflect only twenty-seven different consonantal phonemes; for three of the 30 letters have been tacked on to an earlier 27 without adding any new consonants. Instead of one alef there are three depending on the vowel that follows it; and the 30th letter (*š*) duplicates the 19th (*s*) phonetically, for calendrical reasons. It is remarkable that the ABC which has come down to us had multiple uses, such as numerical and calendrical (to keep track of the days in a lunar month) as well as phonetic spelling. There was also a shorter ABC in which the falling together of consonants yielded the 22-letter ABC of the Arameans/Hebrews/Phoenicians in which some of the Ugaritic tablets are written. And, especially in lexical tablets, Ugaritic words

AUTHOR'S NOTE: The references are (except where stated otherwise) from *UT* (Gordon 1967), because it is the only extensive corpus of texts with a detailed grammar and full glossary. *KTU* (Dietrich, Loretz, & Sanmartín 1976) is valuable insofar as it keeps adding the new texts. However, like *CTA* (Herdner 1963) that preceded it, *KTU* has changed the text numbers, injecting chaos in a field that instead needs order. The system of numbering in *UT* was designed to accommodate future discoveries and at the same time to facilitate checking with the *editio princeps*. This has not been understood by the authors of *CTA* and *KTU*, who apparently regard their disconcerting renumbering of the tablets as part of their scholarly contribution.

The spate of important publications on Ugarit goes on unabated. The most comprehensive and useful is Pardee 1988.

are sometimes inscribed in the Mesopotamian syllabary. And finally we may note that individual scribes have written tablets which confront us with phonetic solecisms.

The alef (א) appears in three different forms which are transliterated *a*, *i*, and *u* respectively. Their normalization, however, calls for a consonantal alef: [ʔa], [ʔi], [ʔu]. The vowel inherent in an alef letter reflects the vowel following the alef; in the case of a vowelless alef, it usually reflects the preceding vowel.

Semitic [b], [g], and [d] normally come into Ugaritic unchanged as *b*, *g*, and *d*. However, Semitic *b* may interchange with *p* in the same word; thus *šbh* and *šph* 'family' occur (Heb. משפחה) and similarly *nbk* and *npg*.

The phoneme [d] is preserved only sporadically in Ugaritic (i.e., represented by the letter *d*); normally it shifts to [d] as in standard Syro-Aramaic.

Semitic [h], [w], and [z] come into Ugaritic unchanged and are written *h*, *w*, and *z*.

Semitic [ḥ] (which falls together with [h] in Hebrew and Aramaic) is preserved in Ugaritic and written *ḥ*. It should also be observed that one of the words for 'young man' in Ugaritic is *ḥrd*. (That it is not a loanword in Ugaritic is indicated by its presence in Eblaite, a millennium before Ugaritic.) The word is common in Egyptian, where it is written *ḥrd* (with *ḥ* rather than *ḥ*). Since Ugaritic is Egypto-Semitic, we may accordingly note that Egypto-Semitic [ḥ] converges with [h] in Semitic.

Semitic [t], [y], [k], [l], [m], [n], [s] and [ʔ] come into Ugaritic unchanged and are written with the letters *t*, *y*, *k*, *l*, *m*, *n*, *s*, and *ʔ*.

Semitic [ǧ] is preserved in Ugaritic and is represented by the letter *ǧ*. However, there are several words in which Arabic [z] corresponds to *ǧ* in Ugaritic. Ordinarily, Arabic *z* corresponds to *z* in Ugaritic. It has been tentatively suggested that the correspondence of Ugaritic *ǧ* with Arabic *z* may reflect a hitherto unknown Semitic phoneme.

Semitic [p] comes into Ugaritic unchanged and is indicated by the letter *p*. But occasionally, as noted above, *p* and *b* are interchanged in variant spellings of the same word.

When we find [s] in all the other Semitic languages, it is also written with the letter *s* in Ugaritic. Usually Semitic [d] falls together with [s] and appears as the letter *s* in Ugaritic. However, in text 75, Semitic [d] appears as the letter *z*. Dialectically, classical Arabic *d* is pronounced *z*; e.g., in Iraq.

In text 77, Semitic [t] appears as the letter *z*.

Usually Ugaritic *q* corresponds to Semitic [q]. However, Ugaritic *q* interchanges with Ugaritic *g* in the variants *q̄t* (Krt:223) 'bellowing' // *q̄t*

(Krt:120) (Heb. שָׁאָגה 'roar'). (In Bedouin and in Iraqi Arabic, [q] is commonly pronounced [g].)

Semitic [r] remains *r* in Ugaritic but final [-r] may be dropped. The name of the god Kōṭar appears as *kt in* ('nt:VI:18) and the jussive *yatbur* is written *yṭb* (1 Aqht:108, 123) 'may he break', as shown in duplicate passages (*yṭbr* in lines 137, 149).

Both [š] and [ś] fall together as *š* in Ugaritic. The same thing happens in unpointed Heb. and yet ש [ś] and ש [š] have been sharply distinguished since Masoretic times down to the present. The causative prefix (š-) of the Shaf'el conjugation shifts (by assimilation) to [t] in roots where the first consonant is [t].

Semitic [t] regularly comes into Ugaritic as *t*.

The vowels inherent in the three alefs do not indicate that any of the three main Semitic vowels [a, i, u], long or short, have been altered or dropped in Ugaritic in any position, except for (1) the conditioned shift of [a?] to [e?] at the end of a syllable, and (2) vowel harmony.

The shift of [a?] to [e?] is illustrated by words like **[raʔš-]* 'head' (Arabic [raʔs-]), which developed from **[raʔš-]* into **[rā(ʔ)š-]* and finally with the Canaanite shift of accented long *ā* to *ō* into Hebrew [rō(ʔ)š-]. But in Ugaritic it is [reʔš-] (written *riš*); cf. Akkadian and Aramaic [rēš-] 'head'. Similarly, Semitic **[ḏaʔn-]* 'small cattle' = 'sheep and goats' appears as [šēn-] in Ugaritic (but as [šō(ʔ)n] in Hebrew). Note that this shift cuts across East and West Semitic. It appears selectively in Hebrew, starting with the first word in the Old Testament: (*bē*)-*rē*(ʔ)*šīt* < **[raʔšīt]*.

The Canaanite shift of long *ā* to *ō* does not take place in Ugaritic. Thus the feminine plural suffix [-āt] does not shift to [-ōt]; *ksat* 'thrones' (vs. Hebrew כְּסָאוֹת). Similarly, *šmal* [šimʔāl-] 'left' (vs. Hebrew שְׂמָאל). The preservation of the *ā* cannot be attributed solely to the early date for Ugaritic, for the shift to *ō* is already attested in Canaanite words spelled syllabically in the Amarna tablets.

Vowel harmony is apparent in Ugaritic words such as *ʕllūp* 'prince, chief' (Hebrew אֲלִיָּה), *udm* 'Edom' (Hebrew אֶדוֹם), *urbt* 'window' (Hebrew אֲרָבָה), etc. For a suggested formulation of the conditions under which vowel harmony takes place in Ugaritic, see *UT* § 5.19.

The final *-m* of mimation is dropped except in a few adverbial accusatives such as *gm* 'aloud (lit. 'with voice')' and *bkm* 'weepingly, tearfully'.

Initial **[w-]* shifts to [y-] in Ugaritic, as in Northwest Semitic (Canaanite and Aramaic). The *ʕ* verbs provide many examples; e.g., *ybl* 'to bring', *ydd* 'to love', *yld* 'to bear (a child)', *yṣ* 'to go out', *yrd* 'to go down', *yṭb* 'to sit'.

Note also *yrḥ* (9:11) [yarḥ-] ‘month’ (Hebrew יָרֵחַ) and *yrḥ* (77:4) ‘(the moon-god) Yariḥ-’ = Hebrew יָרֵחַ ‘moon’ (vs. Akkadian [warḥ-] ‘month’).

The assimilation of [n] to a directly following consonant is normal in Ugaritic. Thus *at* for masculine [ʔatta < *ʔanta] or feminine [ʔatti < *ʔanti] ‘thou’, *mšb* [maššab- < *maṣṣab-] ‘stand or beam (of scales)’, *gt* [*gint-] ‘wine or oil press’ like *bt* [*bint-] ‘daughter’. This assimilation does not take place with third radical [n] in verbal forms like *ytnt*, *mgntm*. Cf. all Hebrew final-*n* verbs, except *ntn* in which the final radical is (unlike the [n] in Ugaritic *ytnt*) assimilated; thus Hebrew נָתַתָּ [natatta] ‘thou hast given’. Note that the [-n] of the preposition *min* ‘from’ (which is exceedingly rare in Ugaritic) is assimilated to the following alef in *mab* (1015:11) = מֵאֵב ‘from father’.

The [l-] of *lqh* ‘to take’ is assimilated to the directly following [q], as in *iqh* ‘I shall take’ (like Hebrew אָקַח).

In reduplicated biconsonantals, a final radical [n] may or may not be assimilated. Note *knkn* beside *kkn(t)*.

As in Hebrew, so too in Ugaritic, [-nh-] may or may not be assimilated to [-nn-]. Thus ‘I shall bury him’ appears interchangeably as *aqbrn* (1 Aqht:126) or *aqbrmh* (1 Aqht:111) in otherwise identical contexts. Cf. the accusative suffix in Hebrew ‘he will guard him’ with assimilation (יִשְׁמְרֵנִי) or without it (יִשְׁמְרֵנִי).

The diphthongs [ay] (certainly) and [aw] (probably) are at least sometimes reduced to [a] instead of to [ē] and [ō], respectively. The presence of the reduction of [ay] to [a] in Ugaritic was demonstrated in an unusual and unexpected way. A Minoan wine jar from Knossos was labeled by the Linear A syllables *ya-na*, which I identified as meaning ‘wine’ even though in Northwest Semitic the word was then expected to appear as [yayn-] or [yēn-]. Subsequently, the name of a village in the Ugaritic realm was brought into the discussion. The village is spelled *yn* in the Ugaritic alphabet, but in Akkadian texts from Ugarit it is written *ā^lya-na* or *ā^lGEŠTIN^{na}*. The latter spellings show that the name *Ya-na* means GEŠTIN ‘wine’ in Ugaritic. Accordingly, the diphthong [ay] has been reduced, not to [ē], but to [a]. This has many broad ramifications. To cite only one: in Hebrew inscriptions such as the Samaria ostraca the word for ‘wine’ is written *yn*, which, for all we know, is to be normalized [yan] rather than [yēn]. Hitherto, all Semitists regarded [yēn] as the only possibility (see Rendsburg 1990).

The voiced dental stop [d] is assimilated to the corresponding emphatic [t] in the personal name *Šdqšlm* (under the influence of the emphatics [š] and [q]) by the scribe of text 1005:4, 10, 14, who spells the name *Štqšlm*.

The second [ʕ] in the personal name **ʕdʕnt* ('Servant of the goddess 'Anat') is dissimilated to zero in the atypical tablet 1045:4, 8, where the name is written *ʕdnt* [ʕabdanat] < *[ʕabd-ʕanat].

The dissimilation of [m – m] to [l – m] takes place in *lḥmd* [laḥmad] (51:V:101) for *mḥmd* [maḥmad] (lines 78, 94) 'pleasantness, choiceness' = 'the best of ...'.

The [t] of *tṭb* '(the god) Teshub' is partially assimilated to the voiced [d] of the preceding [g] in the Hurrian personal name *agdṭb* [ag(i)deṭub].

The scribe of the atypical text 1045 spells the names usually written *tlmyn* as *ṭlmyn* (with emphatic [ṭ] for unvoiced [t]) in 1045:7; and conversely spells the personal name *ypltn* as *ypltn* in 1045:4.

Thrice in 1045, voiced [ḡ] occurs for unvoiced [ḥ]; to wit, in the personal names *ʕdyrg* (:2), *ḡyrn* (:3), and *ṣḡr* (:13), which are elsewhere spelled *ʕdyrh*, *hyrn* and *shr*. Note that [r] occurs in all three, and also that the letter [ḥ] does not appear anywhere in 1045.

The voiced [d] of the root *ṣdq* 'righteous' in the personal name *ṣdqn* is changed to voiceless [t], atypically, in *ṣtqn* (in 1153:2, 3; 1154:4, 6, 7). Moreover, the emphatic [q] is dissimilated to unvoiced [k] atypically in *ṣdkn* (1045:6) under the influence of the emphatic [ṣ].

The labials [b] (voiced stop) and [m] (bilabial nasal) occasionally interchange. Thus *bbqr* (Krt:113) = *bmqr* (Krt:216f.) 'in the well', where the [m] has been assimilated to the preceding [b]. Conversely, the [b] of *ybm̄t* in Anat's familiar epithet *ybm̄t . limm* (. represents the word divider character) appears as *ym̄mt . limm* in 'nt:III:9. It is interesting to compare the personal name *מִימָה*, Job's second daughter, in English *Jemimah*.

As in Hebrew *אחת*, the feminine of the numeral '1' also in Ugaritic is *aḥt* (i.e., masc. *aḥd* + fem. -t) [ʔaḥ(h)att- < *ʔaḥ(h)adt-] with [-dt-] assimilated to [-tt-]. The same assimilation takes place in the conjugation of verbs where the third radical [d] is followed by a suffix beginning with [t]; e.g., *ylt* [yalattā] 'they (3rd fem. dual) have borne' from *yld*.

The loss of the 1st common sing. preformative [ʔa-] before [ʕ] in sandhi: *ʕbk* (3 Aqht:'obv':22) 'I shall set thee'; *wank . ny* (137:28) [wa-ʔanakufniyu] 'and I shall answer'. Note that this phenomenon is indicated in the Masoretic treatment of פֿע verbs (where the פֿ is historic—not phonetic: *וַאֲנִי* (1Kgs. 11:39), *וַאֲנִי* (Zech. 11:5)).

Note [tt] for [t] in *tṭl* 'dew' (1 Aqht:200); elsewhere it regularly is written *tl*. That this is not a scribal error is indicated by the name of Baal's daughter *Tly*, which is spelled *Tṭly* in 67:V:11.

Hkpt ‘Aigyptos/Egypt’ (originally the designation the Memphis area, now the Cairo area) is once written *hqkpt* (‘nt:VI:13) reflecting a vacillation between representing the palatal stop as *k* or *q*. Note that the Old Testament word for ‘helmet’ (in modern Hebrew it means ‘hat’) is written either קובע or כרבע.

References

- Dietrich, Manfred; Oswald Loretz; & J. Sanmartín. 1976. *Die keilalphabetischen Texte aus Ugarit einschließlich der keilalphabetischen Texte außerhalb Ugarit*, part 1. *Transkription*. Kevelaer: Butzon & Bercker; Neukirchen-Vluyn: Neukirchener Verlag (Alter Orient und Altes Testament 24).
- Gordon, Cyrus H. 1967. *Ugaritic Textbook*. Rome: Pontifical Biblical Institute (Analecta Orientalia 38).
- Herdner, Andrée. 1963. *Corpus des tablettes en cunéiformes alphabétiques découvertes à Ras Shamra–Ugarit de 1929 à 1939*. Paris: Imprimerie Nationale; Geuthner (Bibliothèque archéologique et historique 79, Mission de Ras Shamra 10).
- Pardee, Dennis. 1988. *Les textes para-mythologiques de la 24e campagne (1961)*. Paris: Éditions Recherche sur les Civilisations (Mémoires 77, Ras Shamra–Ougarit 4).
- Rendsburg, Gary. 1990. “Monophthongization of *aw/ay* > *ā* in Eblaite and in Northwest Semitic.” *Eblaitica* 2: 91–126.

Phoenician and Punic Phonology

Stanislav Segert

University of California, Los Angeles

4.1. Terms and problems

The term “phonology” can be applied to Phoenician and Punic materials in its most narrow sense: dealing with reconstructed phonemes and their relationships.

Phonological evidence for Phoenician is preserved only in written records from antiquity, in various scripts. Data and also analogies from other languages have to be used for reconstruction and characterization of Phoenician and Punic phonemes. These attempts to introduce necessary phonetic criteria are in many respects uncertain, due to both the variety and inconsistency of the written sources, to dialectal differences and developments not clearly attested, and to an assorted range of problems having to do with the cognate languages.

It is possible to establish a fairly uniform phonological system for Phoenician used in its older period, during the first half of the first millennium B.C., at the eastern shore of the Mediterranean Sea. But the later extension of colonies toward the west, and then the influence of other languages, some of them non-Semitic, led to various changes. Some of them are difficult to trace, due to uncertainties and inconsistencies in various graphemic systems used for recording later Phoenician and Punic words.

Phoenicians are credited with the invention of the alphabetic script, in which phonemes are consistently indicated by appropriate graphemes. As this system was limited to indicating only consonants in the Phoenician period, it is necessary to search for the data about vowels in records in other scripts and in analogical phenomena in cognate Semitic languages. In Punic gradually some vowels became indicated by consonant letters, but this incomplete evidence has to be supplied by information from sources in other scripts.

AUTHOR'S NOTE: Abbreviations: DN, divine name; NL, place name; PN, personal name.

While for source materials presented below, an exact rendering from the preserved documents is presented, many reconstructions should be provided with a question mark due to circumstances mentioned above.

4.2. Sources

The number of Phoenician texts accessible now is about ten thousand, but most of them repeat one votive formula, with various personal names. As the number of connected texts is very limited, a large amount of data is contained in personal names, of which more than one thousand are preserved.

These texts are written in the Phoenician alphabet, which contains 22 letters for consonants. In Punic, especially in later texts, inconsistencies in expressing consonantal phonemes may be observed.

In later Punic inscriptions some consonant graphemes were used for the indicating of several vowels. A system was developed on this basis.

Another system imitated the model of Roman vowel letters, which was itself based on Phoenician letters, accepted via Greek and Etruscan intermediation. E.g. ' = *a*: *t'ḥt* /*taḥt*/; *h* = *e*: *bḥrm* /*bērīm*/.

Punic conversation is recorded in Roman letters in the comedy *Poenulus* by Plautus. Both consonants and vowels are rendered in these samples.

As Phoenician was used in many countries over a period of many centuries, various dialects developed.

In the Eastern Mediterranean, the Old Byblian dialect is attested in inscriptions from about 1000 B.C. More inscriptions are preserved from the later periods of the first millennium B.C.; most of them were found in the Phoenician cities on the eastern Mediterranean shore, some farther to the North, in Syria and Cilicia.

Specific dialectal features can be observed in the texts from the Phoenician colonies on Cyprus.

Phoenician texts from the western Mediterranean written during the first half of the first millennium B.C. follow their eastern models.

Some changes appear in the texts from Carthage and other colonies in the second half of the first millennium B.C.; their dialect is called Punic.

Texts written after the destruction of Carthage in 146 B.C. are called Late Punic. They were exposed to the influence of Latin and Berber languages.

In Egyptian texts some Phoenician words, mostly names, are attested; their rendering in the Egyptian mostly consonantal script does not always exactly express the Phoenician consonants; cf. e.g. *k-p-n* for /*g-b-l*/ 'Byblos'.

Babylonian and Assyrian texts in syllabic cuneiform do not render all the consonants exactly enough, but are valuable for expressing all vowels of Phoenician words, especially names.

Similarly the Cypriot syllabic script indicates Phoenician vowels within the Greek context.

Consonantal writing of Phoenician names in the Hebrew Bible corresponds to Phoenician practice, but the indication of vowels can reflect later traditions.

Phoenician names and words are contained in ancient Greek inscriptions from various areas and also in works by ancient Greek authors.

Many Punic names are preserved in Latin inscriptions from North Africa. Punic names and words appear in books by Roman authors.

This variety of sources, different in script type and manner of preservation, is reflected in the system of transliteration and transcription used here. As a matter of principle, only words from Latin sources are rendered by usual Roman letters, while all other scripts are rendered by italics. In transcriptions from syllabic scripts, cuneiform and Cypriot, syllables are separated by hyphens. Words from ancient Greek and Roman inscriptions are given in capital letters, those quoted from literary works transmitted by copying in lower case letters. Reconstructed words are given in Roman letters within slant brackets.

Phoenician: italics, with further indication of the dialect: O(ld)
By(blian); Ph(oenician) in Western Asia; Cy(prus dialect);
Pu(nic), in North Africa; L(ate) Pu(nic), after 146 B.C. E.g. *ḥrm*
OBy; *špnḇʾ* Pu.

Hebrew: *ḥīrām* and *ḥyrwm* (H) /ḥīrōm/

Egyptian: *k-p-n-* (Eg), for Ph /gubl-/ NL ‘Byblos’

Babylonian and Assyrian: *ḥi-ru-um-mu* (cuneiform syllabary) for
/ḥīrōm/

Cypriote syllabary: *mi-li-ki-ya-no-to-se* (Cy syll.), for /milki-/

Greek epigraphic: *ThENNEITH* (Gr), DN ‘Tanit’

Greek text: *amilchar*

Latin epigraphic: ANNIBAL

Latin text: hannibal

Reconstructed forms: /ḥīrōm/

Hypothetic forms: **ʾaḥat-milk*

4.3. Kinds of phonological data

While all direct evidence about Phoenician phonological phenomena is contained in texts written in antiquity, for phonological reconstructions phonetic observations of languages preserved from antiquity by oral tradition, such as Hebrew, or even modern languages, such as Arabic, are necessary.

Knowledge of ancient phonological features is limited to the signs and graphical devices used in the ancient scripts.

The evidence about consonants is adequately provided by the Phoenician alphabet of 22 consonant graphemes. In the cuneiform syllabary some consonants are not clearly distinguished. Not all Phoenician or Punic consonantal phonemes can be directly rendered in the Greek and Roman alphabets.

No vowels are indicated in Phoenician inscriptions. The indication of vowels in Punic texts by consonant letters is not always complete and often inconsistent. Special vocalic signs in the Phoenician names quoted in the Hebrew Bible reflect a later tradition of pronunciation. Vocalic elements of cuneiform syllabic signs indicate Phoenician vowels, but not always exactly. There is no specific manner to distinguish /o/; it is indicated by the same sign element as /u/. The Greek alphabet accurately expresses Phoenician vowels. In the Latin alphabet the basic vocalic values are indicated.

For the quantity of vowels there are no specific signs in the Phoenician alphabet. In some Punic texts some long vowels are marked by phonologically related consonant signs. Reduced vowels are not indicated as such in the Phoenician script. Long /ē/ and /ō/ can be clearly indicated in the Greek alphabet, while the reduced vowels may be suggested by *y*. The length of vowels is sometimes indicated by repetition of vocalic elements in the syllabic cuneiform.

While doubling of consonants is not indicated in Phoenician script, it is marked by the repetition of Greek and Roman letters and sometimes of consonantal elements in syllabic cuneiform.

In no ancient script used for recording Phoenician words are graphical signs for word stress attested. Accent marks in some Greek manuscripts were added later.

The use of analogy cannot be avoided, due to the incomplete attestation of Phoenician in the ancient written texts.

The close relationship of Phoenician to another Canaanite language, Hebrew, makes it possible to use the phonetic tradition transmitted in the reading of the Hebrew Bible. For consonants this comparison is reliable; for vowels it is necessary to take into consideration both the specific differences

between Hebrew and Phoenician and the relatively late time for the fixation of the Hebrew vowel signs.

Another Canaanite language, Ugaritic, may be used for some reconstructions. This archaic language from the late Bronze age, before 1200 B.C., was written in an alphabetic cuneiform script in which some vowels were indicated.

Both literary Arabic and the contemporary Arabic dialects of Lebanon and Syria can be used as help for a better understanding of Phoenician phonology. The basic values of consonants in modern Semitic languages are based on phonetic recordings and observations of the Arabic pronunciations.

4.4. Consonantal phonemes: Inventory and systemic changes

The Phoenician consonant inventory (Table 4-1) can be reconstructed with help of Hebrew and Arabic. The partial disintegration of this system in Punic does not allow presentation of a corresponding table.

Table 4-1. Phoenician Consonants

				Linguals	Nasals	Semivowels
Laryngeals		ʾ	h			
Pharyngeals		ʿ	ħ			
Velars	q	g	k			
Palatals			š			y
Sibilants	š	z	s	l		
Dentals	ṭ	d	t	r	n	
Labials		b	p		m	w
Emphatic				Voiced	Unvoiced	

It is possible that Phoenician in its oldest period, at the beginning of the Iron age about 1200 B.C., had more consonant phonemes (as had Ugaritic), but inscriptions on arrowheads do not provide sufficient evidence to prove this.

The weakening and elision of some consonants and the apparent confusion of some consonant signs were influenced and accelerated by the close

contact of Punic in North Africa with non-Semitic languages, viz., Berber and Latin.

The laryngeal /ʔ/, from phonetic viewpoint glottal stop, was no longer used in writing to indicate a specific phoneme; e.g. *mlkt*—instead of original *mlʔkt*—‘work’.

Weakening of the pharyngeal /ʕ/ can be observed: *šm* ‘he heard’, then *šm*; *SAMŌ*.

Laryngeal /h/ and pharyngeal /ħ/ seem to have persisted relatively longer. They could be indicated by the Latin letter *h*. For /h/ cf. *mhrbl*, maharbal (PN), but a-eliḥot ‘the hospitality’, *ha-he-; for /ħ/ cf. *ḥnbʿl*, hannibal, but also ANNIBAL; use of letter for weakened voiced pharyngeal ʕ: *ʕm* for older *ḥym* ‘life’. (Names are transliterated from Greek and Latin without capital letters at the beginning, in accordance with the ancient writing practice.)

Oppositions between voiced and unvoiced and between emphatic and non-emphatic consonants were disregarded, especially in Punic, more frequently in its later stages. This confusion affected mostly sibilants. E.g. *mls* instead of *mlš* ‘interpret’; demonstrative pronouns *s*, *st*, *sith* ‘this’ instead of *z-*; *mšl* instead of *mzl* ‘fate’.

4.5. Vocalic phonemes: Inventory and systemic changes

Unlike for consonants, the evidence for Phoenician vowels is scanty. Data from words, mostly in not always adequate non-Phoenician writing systems, have to be supplemented by often hypothetic information from the corresponding features in related Semitic languages.

An attempt to reconstruct the Phoenician vocalic system is in Table 4-2.

Table 4-2. Phoenician Vocalic System

Short vowels	i	a	u
Long vowels	ī	ō	ū
Long vowels	ē (<*ai)	ō (<*au)	

This system corresponds with that of the archaic Canaanite language, Ugaritic, and with that of the conservative Classical Arabic.

The monophthongization of diphthongs occurred early, as no *y* or *w* letters used in related languages for the indication of diphthongs are attested. Cf. *bt*/bēt/ ‘house’, Greek name of letter *bēta* (but *baitylon* in Sanchuniaton); *mōt* ‘death’.

The Canaanite vowel shift of **ā > /ō/* can be explained as conditioned by phonetic behavior: if the mouth is not sufficiently opened, pure long *ā* cannot be produced, the sound tends toward *ō*. Cf. PN /ḥīrōm/, *ḥi-ru-um-ma*, *ḥyrwm* and *ḥīrōm* (Hebrew), *eirōmos*; /-yatōn/ ‘he gave’ in PNN *mi-li-ki-ya-to-no-se* (Cypriot), BALIATON; /-malōk/ ‘he reigned’ in PN *ba-al-ma-lu-ka*; NADŌR ‘he vowed’; /-milkōt/ ‘queen’ in PN *ab-di-mil-ku-ut-ti*.

It seems that this Canaanite change **ā > ō* affected not only the originally long *ā*-vowels, but also vowels lengthened secondarily through word stress: **milkāt > *milkāt > /milkōt/*; perfects like **yatán > *yatān > /yatōn/*; **’adām > *’adām > ADOM* ‘man’.

The change in this direction continued in Phoenician (unlike in Hebrew) all the way toward /ū/. It affected both *ō*-vowels originated by monophthongization from **au*; e.g. **qaul > *qōl > /qūl/*, KOULŌ ‘his voice’—and those which resulted from original **ā*: ADOUN ‘lord’ (cf. Hebrew *’ādōn < *-ān*), salus /šalūš/ ‘three’, SANUTH /šanūt/ ‘years’.

The vowel /e/ in other than Phoenician script can be considered an allophone of /i/; cf. *ers* and *chirs* ‘shard’. This may be valid also for some instances of /ē/: *abdēlimos* PN ‘Servant of gods’. (‘Servant’ with capital S, since in English proper names are written so.) Similarly /o/ is related as allophone to /u/; cf. MOTTHUN and MUTTHUN PN ‘Gift’.

For vowels outside of the system the letter *y* could be used in Greek and subsequently in Latin texts: for **ū*: cf. *chyl* and *chil* ‘all’, probably /kūl/; for reduced vowel: *ys, ys, sy* ‘which’, *by-marob* /b^e-/ ‘by protection’.

4.6. Conditioned consonant changes

These changes were caused either by position of the consonant within the word or the syllable or by contact with other phonemes, consonantal or vocalic.

In later Punic, consonants articulated in the dental area, /d/, /t/, /l/, could be eliminated from the end of a word or a syllable. Cf. *ḥmlkt* ~ HIMILCO PN ‘Son of the queen’, *mlk* ‘mr, MOLCHOMOR ~ MOCHOMOR ‘offering of lamb’.

The consonant /y/ at the beginning of a word or its element could be submitted to elision: *yqdš* ~ ‘yqdš ‘he consecrated’, *mlkytn* ~ *mlktn*, MILCATON PN ‘Milk gave’.

The semivowel /y/ was elided between vowels; cf. *bny* /banaya/ in Old Byblian and Phoenician *bn* /banā/ ‘he built’; *ybrky*’/-kiya/ and Punic *ybrk*’/-kā/ ‘may they bless her’.

New consonants can be inserted to avoid hiatus between vowels; cf. the later form of the suffixed pronoun of the 3rd person pl. masc. *-nm* /-nom/ against older *-m* /-om/, e.g. *ḥbrnm* /-ēnom/ ‘their colleagues’ (against BUNOM ‘their son’). An additional consonant /m/ appears in late Punic: *bnm*, BINIM ‘his son’.

Regressive assimilation of /n/ is amply attested in Phoenician; e.g. *št* /šatt-/ < **šant-* ‘year’, *kt* /kattī/ ‘I was’ from the verb /k-w-n/. In later Punic, however, forms with non-assimilated /n/ reappear, e.g. *mnšbt* beside *mšbt* ‘stela’.

Some unusual forms can be explained by dissimilative tendencies. Cf. *n’spt*, against *m’spt* ‘assembly’, by the dissimilation of labials at a distance.

The consonant /h/ can be assimilated to a preceding or following /i/ into /y/. Cf. Old Byblian *ḥh* /-ihu/ and Cypriot *ḥy* /-iyu/ ‘his father’.

4.7. Conditioned vocalic changes

Since most of the evidence for vowels and their changes is based on records in scripts which did not render them consistently, the presentation is hypothetical.

Changes caused by word stress are discussed below. Short **ā* was lengthened by the impact of stress into **ā̃* which was changed into /ō̃/, through a systemic change, the Canaanite vowel shift.

Even as there are no specific signs for reduced vowels in the Phoenician script, they can be traced if they are rendered by the Greek and subsequently Roman letter *y* or if there is uncertainty in their rendering. For examples see § 4.5, also *bynuthi* ‘my daughters’. Variety of forms in Greek letters of the plural in construct state ‘face(s) (of)’, *PhANE* and *PhENĒ* can point to a reduced vowel in the first syllable. The quality of the original vowel before its reduction cannot always be traced with certainty.

Vowels in the initial syllable were often omitted after the glottal stop /ʔ/ preceding them was no longer respected. Cf. *ḥrm* < *ḥrm* /ʔaḥī-/ PN ‘My brother ...’, *bbʔ* and *ʔbbʔ* PN ‘My father ...’, *donni* ‘my lord’ ~ *ʔdn*; in particles: *t*, *Th* (originally *ʔt*) nota accusativi introducing object, perhaps *p* (cf. *p*) ‘also’.

4.7.1. The development of secondary vowels

Prothetic vowels indicated in writing by ʔ (ʔaleph): *ʔršp* /aršap/, cf. *ršp* DN; Cypriot *ʔz* (cf. *z*) ‘this’, YMU ‘(what) > which’; before a preposition: *ʔbbt* /ʔab-bēt/ ‘in house’.

Anaptyctic vowels: in segolate nouns: *qb'r* /qabar/ (cf. *qbr* /qabr/) 'grave', *s'r* 'ten', *syris* (cf. *šrš*) 'root'.

Assimilation to vowels: u-ulech < *(h)a-(h)ū- 'the visitor'; perhaps *ui* < **uhuya* 'my brother'.

Assimilation to consonants: to labials, in direction to /u/: *moutinas*, MUT-TUN, muttines (cf. *mattēn*, MITUN, *ma-ta-an*-) PN 'Gift'; perhaps *BYN*, byn /bün/ 'son' (cf. Ugaritic *bun*-).

4.8. Other phonological features

Only one suprasegmental feature can be observed, word stress, but only by indirect suppositions. Syllabic structures have to be mentioned in a survey of phonology.

No graphic signs or direct information about word stress is available. But by observation of the change **á* > **ā* > /ō/ the position of word stress can be determined on the last syllable in most instances.

Open syllables appear both with short (CV) and long vowels (C $\bar{V}$). Closed syllables had probably only short vowels in the center (CVC). Also closed syllables with the cluster of two consonants on the end can be observed (CVCC): OTMILC < **ahát-milk* PN fem. 'Sister of the king'; *qrthdšt* /qart-ḥadašt/ NL 'New City', cf. derived adjectives carthadati and *KARKhA-DONION*, cf. Latin CARTHAGO and Greek *karchēdōn* from shortened form **karthadō*.

For the dropping of syllables in the middle of a word, cf. BONCAR and BODMILKAR, *bdmlqrt* PN 'In hand of Melqart'.

Selective Bibliography

- Benz, Frank L. 1972. *Personal Names in the Phoenician and Punic Inscriptions*. Rome: Biblical Institute Press.
- Dothan, Aron 1976. "Stress Position and Vowel Shift in Phoenician and Punic." *Israel Oriental Studies* 6: 71–121.
- Friedrich, Johannes, and Wolfgang Röllig. 1970. *Phönizisch-punische Grammatik*, 2nd ed. Rome: Pontificium Institutum Biblicum.
- Fuentes Estañol, María José. 1980. *Vocabulario Fenicio*. Barcelona: Biblioteca Fenica.
- Harris, Zellig S. 1936. *A Grammar of the Phoenician Language*. New Haven: American Oriental Society.

- Jean, Charles-F., and Jacob Hoftijzer. 1965. *Dictionnaire des inscriptions sémitiques de l'Ouest*. Leiden: Brill.
- Moscatti, Sabatino, ed. 1988. *The Phoenicians*. New York: Abbeville.
- Segert, Stanislav. 1955. "Zum Übergang $\bar{a}$ > $\bar{o}$ in den kanaanäischen Dialekten." *Archiv Orientalní* 23: 478.
- . 1976. *A Grammar of Phoenician and Punic*. Munich: Beck.
- . forthcoming. "113b, Semitic: Phoenician and Punic." In *Namengebung – Proper Names*. Berlin: de Gruyter.
- . forthcoming. "Phoenician and Eastern Canaanite Languages." In *The Semitic Languages*, ed. Robert Hetzron. London: Routledge.
- Veenhof, K. R. 1973. "Phoenician-Punic." In *A Basic Bibliography for the Study of the Semitic Languages*, ed. J. H. Hospers, vol. 1, pp. 146–71. Leiden: Brill.

Ancient Hebrew Phonology

Gary A. Rendsburg

Cornell University

5.1. Hebrew and the Semitic languages

Hebrew is a Semitic language, attested since ca. 1100 B.C.E. as the language of the Israelites (the Bible also uses the ethnonym Hebrews, and later the term Jews becomes more common). Ancient Hebrew died out as a spoken language in the third century C.E., though it was retained in an unbroken chain for liturgical and literary purposes unto the modern era. In the late 19th and early 20th centuries, Hebrew was revived as a spoken language. It is used today as the national language of Israel. Not surprisingly, during its history of more than three millennia, the language has undergone various changes, especially in the realm of phonology. This chapter is devoted to ancient Hebrew, defined here as the period of ca. 1100 B.C.E. to ca. 250 C.E., with a particular emphasis on historical matters. Occasionally, later developments in the medieval period also will be noted. For the phonology of Modern Hebrew, see Chapter 17.

Semitists continue to debate the classification of the individual Semitic languages, but all agree that Hebrew falls within the Northwest Semitic group. The languages of this group are Amorite, Ugaritic, Canaanite, and Aramaic. According to many scholars (myself included), Ugaritic is to be subsumed under Canaanite, but the former is attested in the second millennium B.C.E. and the latter almost exclusively in the first millennium B.C.E., so for the nonce I distinguish them. An additional Northwest Semitic language may be Eblaite, though a majority of scholars holds that it is more closely linked to Akkadian (East Semitic).

In essence Hebrew is but a dialect of Canaanite. The other dialects of this language are Phoenician, Ammonite, Moabite, Edomite, and Deir ‘Alla (referring to the epigraphic remains found at Tell Deir ‘Alla a few miles east of

AUTHOR’S NOTE: I am grateful to Alan Kaye and Saul Levin for their comments on an earlier version of this essay. We do not agree on all the points raised, and I alone am responsible for the views expressed. I also extend thanks to my graduate students Scott Noegel and Richard Wright for their critical reading.

the Jordan River, though other opinions hold that Deir 'Alla is a dialect of Aramaic or an independent branch of Northwest Semitic altogether). These dialects of Canaanite, attested mainly in the first millennium b.c.e., were all mutually intelligible, and probably were differentiated no more than, say, the geographical varieties of Modern German or Modern English.

Phoenician, Ammonite, Moabite, Edomite, and Deir 'Alla are known primarily through inscriptions found in archaeological excavations in the Levant (Phoenician is an exception in two regards: [a] epigraphic remains have been found throughout the Mediterranean region, and [b] occasional classical writers, especially Plautus, preserve material). The total amount of known material would fill only a slender volume. The corpus of ancient Hebrew, by contrast, is quite large. The sources are the Hebrew Bible (Old Testament), the book of Ben Sira (one of the Apocrypha), the Dead Sea Scrolls found at Qumran, the Mishnah and other works authored by the rabbis of late antiquity, and various inscriptions (some of considerable length, but hundreds are very short, often consisting of only personal names).

Much of the following discussion concerning Hebrew phonology also may hold for the other Canaanite dialects, but our knowledge of these dialects is limited. On the other hand, we know that some of the other varieties of Canaanite were differentiated specifically in the realm of phonology (see the above comparison with German and English dialects, and see below for an occasional point of contrast).

5.2. Variation within Ancient Hebrew

Until now I have spoken of Hebrew as if it were a unified dialect within Canaanite, but this is an oversimplification. In fact, ancient Hebrew may be distinguished in various ways.

A) Based on differences visible in the Bible, diachronically we can distinguish Archaic Biblical Hebrew (ca. 1100–1000 B.C.E.), Standard Biblical Hebrew (ca. 1000–550 B.C.E.), and Late Biblical Hebrew (ca. 550–200 B.C.E.). The Hebrew of the Dead Sea Scrolls, known also as Qumran Hebrew (after Qumran, the site of discovery of these documents), is a continuation of Late Biblical Hebrew, and is attested ca. 200 B.C.E. – ca. 70 C.E.

B) Ancient Hebrew had various regional varieties. This finding also is based on various differences visible in the Hebrew Bible, and is confirmed in some instances by the epigraphic remains. Here we may distinguish Judahite

Hebrew, i.e., the regional dialect used specifically in Judah and its capital of Jerusalem, versus Israelian Hebrew, i.e., the dialect bundle of all other areas of traditional Israelite territory (areas such as Samaria, Galilee, and Transjordan). The vast majority, about 80%, of the Bible is written in Judahite Hebrew, and the remaining sections are written in Israelian Hebrew. I refer to Israelian Hebrew as a dialect bundle, because almost certainly there were minor differences between, for example, Transjordanian Israelian Hebrew and Galilean or Samaritan Israelian Hebrew. The Transjordanian variety no doubt shared many features with Ammonite, Moabite, and Deir 'Alla; while the Galilean variety no doubt shared many features with Phoenician (and with Aramaic too). However, the available data generally do not allow us to isolate such minor differences, and for the most part it suffices to speak of Israelian Hebrew as a unified group of local varieties which, as a whole, contrasts with Judahite Hebrew.

C) Ancient Hebrew also was characterized by diglossia. The Bible, Ben Sira, and the Dead Sea Scrolls are written in the literary standard. But everyday speech differed considerably, as can be determined by occasional departures from the classical norm in these texts, especially when these phenomena parallel colloquial developments known from other spoken varieties of Semitic (e.g., colloquial Arabic). In late antiquity, the colloquial dialect was utilized to record texts such as the Mishnah and related works, so that the term Mishnaic Hebrew is used. The data at our disposal which allow us to posit diglossia in ancient Hebrew are mainly in the realm of morphology. Differences in phonology are more difficult to demonstrate.

In presenting the phonology of ancient Hebrew, in the main we refer to Standard Judahite literary Hebrew, i.e., the literary variety utilized in Judah ca. 1000–586 B.C.E. But where the data permit us to witness distinct usages in other varieties of ancient Hebrew, these will be noted.

5.3. Orthography

The Israelites utilized the 22-letter alphabet typically called the Canaanite alphabet (invented by the Phoenicians, according to the standard view). This alphabet represents only consonants, not vowels. Moreover, ancient Hebrew possessed more than 22 consonantal phonemes, so that some of the graphemes (letters) served double duty.

In the earliest Hebrew orthography, vowels were not indicated at all. According to the standard theory, in time, scribal practice led to the adoption

of three letters, <h>, <w>, and <y>, to indicate final vowels. Eventually, this system was expanded to indicate medial vowels as well, though this practice was not carried out consistently. When <h>, <w>, and <y> are utilized in this fashion, they are called *matres lectionis* or vowel letters (see further § 5.6.4, where another, non-standard view of the vowel letters is presented also).

These problems of both consonants and vowels, but especially the vowels, created a certain ambiguity in the reading of ancient Hebrew. The extent to which such ambiguities caused readers problems cannot be determined, but probably in general usage no undue hardship arose. However, because the biblical books achieved a level of sanctity in Judaism, no amount of ambiguity could be tolerated in the reading of sacred literature. An official reading tradition existed, in which the reader of the Bible (for example, in the synagogue for liturgical purposes) read the text in its traditional manner.

In time, a system of vowel markings and other diacritic marks was developed to record the official reading tradition. The people responsible for this notation system are called the Masoretes (tradents) who were active ca. 850 C.E. My reconstruction of the history here is actually a bit too simplistic; in reality there was more than one official reading tradition (the Jews of Israel had one main tradition, the Jews of Babylonia another, etc.), and the Masoretic activity actually led to different notation systems too. The normative Masoretic system in use among Jews for the past millennium has been the Tiberian one, named for the city of Tiberias (on the Sea of Galilee) where it developed. Our discussion of the phonology will be based on this system.

The question remains as to how accurately the reading tradition of the biblical text and the Masoretic transcription thereof reflects ancient Hebrew. That is to say, the Masoretic Text (that is, the traditional text of the Bible) dates to ca. 850 C.E. and reflects the manner in which Biblical Hebrew was pronounced at that time. But how traditional, i.e., how ancient, was the reading tradition of the readers for the centuries before ca. 850 C.E.? In other words, does the Masoretic Text reflect Hebrew as it was pronounced five hundred years earlier, one thousand years earlier, even fifteen hundred years earlier? In some cases, we can answer this question, but no definitive conclusion can be reached.

Nevertheless, we will base ourselves on the assumption that the readers of the first millennium C.E. were extremely conservative in their biblical reading tradition, and that the Masoretic Text more or less accurately reflects the pronunciation (or at least one pronunciation) of ancient Hebrew in the first millennium B.C.E., i.e., the time of the composition of the biblical books. I

say “more or less” because, among other points, (a) in some instances we know that the Masoretes no longer recognized consonantal phonemes which were distinguished in ancient Hebrew but which merged only later on, and (b) the system of vowels according to the Masoretic notation has an exceedingly large number of allophones, some or many of which may have developed only after the ancient Hebrew period.

The picture presented in the above outline is further complicated by the fact that there exists an important non-Masoretic reading tradition. The Samaritans, who developed as an offshoot of Judaism ca. 400 B.C.E., also possess the first five books of the Hebrew Bible (the Torah or Pentateuch) as canonical. They have an independent reading tradition for their Scripture, but in this essay we refrain from entering into these differences.

5.4. Phonology of the consonants

At least 29 consonantal phonemes are traceable to Proto-Semitic (comparison with other families in the Afroasiatic phylum suggests the possibility of still other phonemes). The most ancient Hebrew attested retained 25 of these; one local variety of Israelian Hebrew retained one other phoneme; and the remaining three phonemes merged with other phonemes (though one cannot discount the possibility that any or all of these three may have been retained in some restricted geographical locale, lack of evidence notwithstanding). As noted above, the Hebrew (Phoenician) alphabet has only 22 signs, so the recovery of the additional three or four phonemes requires special comment (see below for the individual cases).

Below I list the consonantal phonemes of ancient Hebrew, grouped according to place and/or manner of articulation. Transliteration is based on the standard system utilized in Semitics. Where the IPA symbol differs, it is noted as well. I also note the letter of the alphabet used to render each phoneme.

5.4.1. Bilabial plosives

/p/ – פ.

/b/ – ב.

5.4.2. Interdentals

/t/ (IPA [θ]). In virtually all dialects of Hebrew, this phoneme shifted to /s/, indicated by ש. However, in the Hebrew of Transjordan (specifically Gilead),

as well as in the neighboring Canaanite dialect of Ammonite, this phoneme was retained. The evidence for this comes from the famous passage in Judges 12:6 known as the “shibboleth incident.” The story relates how the Gileadites controlled the fords of the Jordan River. When retreating Ephraimites (from Cisjordan) sought to cross, the guards at the fords asked them to pronounce the word *tibbōlet* [tɪbbōlet], which in Hebrew means ‘stream, torrent’, a fitting password for the crossing of the Jordan River. Since most Israelites did not possess this sound in their phonetic inventory, the Ephraimites would say [sibbōlet], thus revealing the fact that they were not Gileadites. (Compare the manner in which various foreign speakers of English [Germans, for example] pronounce English /t/ as [s], or the manner in which Persians and other non-Arab Muslims pronounce Arabic /t/ as [s].) Since standard Hebrew (and the dialect of Canaanite for which the alphabet was invented) did not possess this phoneme, there was no special grapheme for representing this sound. In the passage just mentioned, Judges 12:6, the letter ש = <š> is used.

For the secondary development of /t/ = [θ] as the fricativized form of /t/, see § 5.5.4.

On the two remaining interdentalals of Proto-Semitic, see § 5.4.13.

5.4.3. *Dental plosives*

/t/ – ת.

/d/ – ד.

/t̥/ – a voiceless emphatic dental plosive, indicated by ט. On the nature of the “emphatics,” see § 5.4.14.

5.4.4. *Nasals*

/m/ – מ.

/n/ – נ.

5.4.5. *Rolled*

/r/ – either a rolled dental or a rolled uvular (its exact articulation is unknown), indicated by ר.

5.4.6. *Sibilants*

/s/ – ס.

/z/ – ז.

/s/ – a voiceless emphatic sibilant (according to most opinions it is a fricative, others hold it to be an affricate), indicated by **ס**. On the nature of the “emphatics,” see § 5.4.14.

/ʃ/ (IPA [ʃ]) – **ש**. Since this letter represented more than one sound relatively late in the history of Hebrew, a diacritical mark was added by the Masoretes on the right side to produce the grapheme **שׁ**. See further § 5.5.1.

5.4.7. *Laterals*

/l/ – **ל**.

/ʃ/ (IPA [ɬ]) – **שׁ**. Since this letter represented more than one sound relatively late in the history of Hebrew, a diacritical mark was added by the Masoretes on the left side to produce the grapheme **שׁ**. See further § 5.5.1.

On the one remaining lateral of Proto-Semitic, see § 5.4.13.

5.4.8. *Velar plosives*

/k/ – **כ**.

/g/ – **ג**.

/q/ – a voiceless emphatic velar plosive, indicated by **ק**. On the nature of the “emphatics,” see § 5.4.14.

5.4.9. *Velar fricatives*

/ħ/ (IPA [x]) – **ח**. This sign was also used to represent /h/. We are able to postulate the existence of both phonemes in the ancient period on the basis of transcriptions of Hebrew words (mainly proper names) in the Septuagint (the ancient Greek translation of the Bible) of the Pentateuch (ca. 250 B.C.E.). When Proto-Semitic comparisons indicate that the consonant /ħ/ is present in the Hebrew word, the Septuagint transcription uses χ (see § 5.4.10 for the practice of transcribing /ħ/). For the eventual merger of /ħ/ and /h/, see § 5.5.2. For the secondary development of /k/ = [x] as the fricativized form of /k/, see § 5.5.4.

/g/ (IPA [ɣ]) – **ג**. This sign was also used to represent /ʕ/. We are able to postulate the existence of both phonemes in the ancient period on the basis of transcriptions of Hebrew words (mainly proper names) in the Septuagint of the Pentateuch (ca. 250 B.C.E.). When Proto-Semitic comparisons indicate that the consonant /g/ is present in the Hebrew word, the Septuagint transcription uses γ (see § 5.4.10 for the practice of transcribing /ʕ/). For the eventual merger of /g/ and /ʕ/, see § 5.5.2. For the secondary development of /g/ = [ɣ] as the fricativized form of /g/, see § 5.5.4.

5.4.10. Pharyngeal fricatives

/ħ/ (IPA [ħ]) – 𐤇. This sign was also used to represent /h/. We are able to postulate the existence of both phonemes in the ancient period on the basis of transcriptions of Hebrew words (mainly proper names) in the Septuagint of the Pentateuch (ca. 250 B.C.E.). When Proto-Semitic comparisons indicate that the consonant /ħ/ is present in the Hebrew word, the Septuagint transcription shows no consonant (see § 5.4.9 for the practice of transcribing /ħ/). For the eventual merger of /ħ/ and /h/, see § 5.5.2.

/ʕ/ (IPA [ʕ]) – 𐤊. This sign was also used to represent /g/. We are able to postulate the existence of both phonemes in the ancient period on the basis of transcriptions of Hebrew words (mainly proper names) in the Septuagint of the Pentateuch (ca. 250 B.C.E.). When Proto-Semitic comparisons indicate that the consonant /ʕ/ is present in the Hebrew word, the Septuagint transcription shows no consonant (see § 5.4.9, for the practice of transcribing /g/). For the eventual merger of /g/ and /ʕ/, see § 5.5.2.

5.4.11. Laryngeals

/ʾ/ (IPA [ʔ]) – 𐤀.

/h/ – 𐤇.

5.4.12. Glides (semivowels)

/w/ – 𐤅.

/y/ (IPA [j]) – 𐤆.

5.4.13. The remaining Proto-Semitic phonemes

There are three remaining traceable Proto-Semitic phonemes: /ḏ/ (IPA [ḏ]), /ẓ/ or /ṭ/ (IPA [ḏʰ]), and /ḏ/ [IPA [ṭʰ]]. There is no evidence for the preservation of these sounds in ancient Hebrew. Instead, in most regional dialects of ancient Hebrew, /ḏ/ shifted to /z/ (in some Israelian dialects it shifted to /d/); and both /ẓ/ and /ḏ/ shifted to /š/ (in some Israelian dialects the former shifted to /t/ and the latter shifted to /q/ or later to /ʔ/). At the same time, scholars recognize that any one, two, or three of these phonemes may have been preserved in some locales. But since the Hebrew alphabet does not have special signs to represent these sounds, it is difficult to ascertain if and where such phonemes may have been retained. Were it not for the story in Judges 12:6 (see § 5.4.2), we would not know that Gileadite Hebrew retained the voiceless interdental /ṭ/, so it is conceivable that elsewhere in ancient Hebrew /ḏ/, /ẓ/, and /ḏ/ existed.

5.4.14. *The nature of the emphatics*

The exact nature of the emphatic consonants /t/, /s/, and /q/ cannot be determined. The corresponding consonants in Arabic are velarized/ pharyngealized; in Ethiopic and Modern South Arabian they are glottalized. Most likely the glottalization is the original Proto-Semitic manner of articulation, so that this can be postulated for ancient Hebrew.

5.5. Historical changes in the consonantal phonology

The consonantal phonology described above is correct for Hebrew in its most anciently attested phase. But already in the biblical period there is evidence for various changes, and in the post-biblical period still more changes are evident. These historical developments will be presented here.

5.5.1. *The shift of /š/ to /s/*

In the course of time the voiceless lateral fricative /š/ shifted to a sibilant and merged with /s/. This is indicated by the numerous interchanges between ש and ס in the spelling of ancient Hebrew. This tendency is less acute in the pre-exilic (pre-586 B.C.E.) books of the Bible, but becomes quite common in the exilic and post-exilic (post-586 B.C.E.) books. Thus, we may conclude that the merger of /š/ and /s/ occurred in Late Biblical Hebrew and continued in still later phases of the language. This shift may be the result of Aramaic influence.

In the centuries after the merger occurred, copyists of the Bible remained faithful to the received text. Accordingly, even though /š/ now was pronounced the same as /s/, in the great majority of cases the biblical manuscripts continued to represent this sound with ש. When the Masoretes devised their system of marking all phonetic distinctions in the received text, diacritic marks were invented to distinguish the two sounds represented by ש. With the dot placed over the upper left hand corner, the grapheme ש̣ represented the former lateral fricative /š/, though now pronounced [s]. With the dot placed over the upper right hand corner, the grapheme ש̤ represented /s/.

5.5.2. *Merger of /h/ and /ħ/ and merger of /g/ and /ʕ/*

In ca. 200 B.C.E., the phoneme /h/ merged with the phoneme /ħ/, and the phoneme /g/ merged with the phoneme /ʕ/. This can be determined from the following. In the Septuagint of the Pentateuch, accomplished ca. 250 B.C.E.,

these phonemes all are represented differently in the Greek transcription of proper names and occasional common nouns (see §§ 5.4.9, 5.4.10). But in the Septuagint of the other books of the Bible, which was accomplished several decades or perhaps even a century later, this consistency disappears. Accordingly, we confidently can fix this phonological development to ca. 200 B.C.E.

5.5.3. *Weakening of the pharyngeals and laryngeals*

In the preceding paragraph we observed that ca. 200 B.C.E. the velar fricatives /ħ/ and /ǧ/ merged with the corresponding pharyngeals /ħ/ and /ʕ/. As time passed, there is evidence for an overall weakening of the pronunciation of the pharyngeals and laryngeals. This can be determined from the Masoretic vocalization system which indicates (a) that the consonants /ħ/, /ʕ/, /h/, and /ʔ/ cannot be geminated (this holds for /r/ as well); (b) that they cannot be vocalized with the vowel shwa, but instead require an auxiliary vowel; and (c) that in final position an anaptyctic vowel is required for all except /ʕ/, e.g., /rûħ/ > [rûaħ] ‘wind’.

In time, in certain locales, this process became extreme. Post-biblical writings (e.g., the Talmud) describe situations in which all the pharyngeals and laryngeals merged. The cities which specifically are mentioned in this regard are Beth Shean, Haifa, and Tivon, all in the Lower Galilee region. Presumably this is due to Greek influence (we know, for example, that Greek influence was strong in Beth Shean). One amusing story records how a certain individual requested a particular item, but the storekeeper could not determine whether he desired *ʾimmar* ‘lamb’, *ḥāmār* ‘donkey’, *ḥēmar* ‘wine’, or *ēmar* ‘wool’. These forms are Aramaic, which was the dominant language in the Galilee ca. 300 C.E., but the story no doubt reflects the situation in Hebrew as well. On the other hand, we have the testimony of Jerome (ca. 400 C.E.) that the Jews mocked the Christians for their inability properly to pronounce the pharyngeals and laryngeals. Accordingly, we may conclude that in some communities Jews retained the original pronunciation of the pharyngeals and laryngeals, while in others they were weakly pronounced or disappeared altogether.

5.5.4. *Fricativization (spirantization) of non-emphatic plosives*

At some point in ancient Hebrew, the six non-emphatic plosives: /p/, /b/, /t/, /d/, /k/, /g/, developed a twofold realization. In post-vocalic position they came to be pronounced as fricatives (spirants); otherwise they retained their original plosive character. The corresponding fricative (spirantized) pronun-

ciations are, respectively: /f/, /v/, /t/ (IPA [θ]), /d/ (IPA [ð]), /k/ (IPA [x]), /g/ (IPA [ɣ]). Almost without exception, these sounds are allophones. Only in rare instances, due to other factors, did phonemic differences arise.

Exactly when the fricativization of the non-emphatic plosives in post-vocalic position occurred cannot be determined. According to one theory, it is due to Hurrian influence, in which case it must have occurred quite early (ca. 1000 B.C.E. [?]). However, most scholars date the fricativization of the non-emphatic plosives in post-vocalic position to a later period, say, ca. 400 B.C.E., perhaps under Aramaic influence.

The reader already has noted that several of these allophones are equivalent to other phonemes in the language. For example, /k/ is the same as /ħ/ (both IPA [x]), and /g/ is the same as /ġ/ (both IPA [ɣ]). Assuming, as most scholars do, that the fricativization of /k/ to /k/ [x] and of /g/ to /g/ [ɣ] occurred ca. 400 B.C.E., and that /ħ/ [x] and /ġ/ [ɣ] were distinguished as late as ca. 200 B.C.E. (see § 5.5.2), then we may posit the coexistence for about two centuries of two sets of one phoneme and one allophone each, phonetically identical (or almost identical).

Similarly, the fricativization of /t/ to /t/ may have resulted in another such case, if we assume that at the same time at least one Hebrew dialect retained the original phoneme /t/ (see § 5.4.2).

Clearly these sounds were pronounced by all (?) Jews ca. 850 C.E. when the Tiberian system of the Masorah was developed. In time, however, the ability to pronounce some of these sounds was lost by various Jewish communities, especially those in Europe.

The three sounds which remained most stable were /v/, /k/, and /f/. Among most European Jews, however, /t/ was realized as [s] (compare the shibboleth incident described in § 5.4.2, though there is no direct connection between the two phenomena). In the two remaining cases, /g/ and /d/, fricativization disappeared and /g/ and /d/ were pronounced as [g] and [d] in all environments. On the other hand, Jews in Arab lands retained most if not all of the fricativized allophones into the 20th century. The Jews of Yemen are an example of a community whose pronunciation of Hebrew included the proper realization of all six allophones.

5.5.5. *Velarization of the emphatics*

Above (§ 5.4.14) we discussed the nature of the emphatics, with the conclusion that originally they most likely were glottalized. Because the corresponding consonants in Arabic are velarized/pharyngealized, and because the majority of Jews in the world ca. 1000 lived in an Arabic-speaking milieu

and themselves spoke Arabic as their native language, in time the emphatic consonants in Hebrew became velarized/pharyngealized as well. This pronunciation remains to the present among the Jewish communities of North Africa and the Middle East.

Jews in Europe, on the other hand, lost the ability to pronounce the emphatic consonants altogether. Thus, in time, /t/ > [t], so that it merged with /t/; /q/ > [k], so that it merged with /k/; and /ʁ/ > [ts], a phoneme common in many European languages, e.g., German.

5.6. Phonology of the vowels

The exact pronunciation of the vowels of ancient Hebrew cannot be recovered. However, we may assume that the classical pattern of Semitic (illustrated best in Classical Arabic) was operative in Hebrew in its earliest historical period. Thus we can reconstruct three basic vowels, either short or long: /a/, /i/, /u/, /â/, /î/, /û/. I utilize herein the circumflex to indicate long vowels which are “pure long” or “etymologically long”—that is, they correspond to long vowels in cognates. By contrast, the macron will be used in the transliteration scheme to indicate short vowels which have been lengthened due to stress—that is, they are “tone long” vowels (see § 5.6.2).

The Masoretic notation system, as noted above (§ 5.3), dates to ca. 850 C.E., and most accurately reflects the pronunciation of Hebrew in the early medieval period. By this time, the classic triangular vowel system had broken down, and numerous allophones had developed, based on a complex system of syllabification and accentuation. Again, exactly when the shift from the basic three vowels, short or long, to the system to be described below occurred, is unknown. But it is apposite to quote the view of Jerome (ca. 400 C.E.): “It is of no consequence whether [the word Shalem] is pronounced Salem or Salim, because Hebrew very rarely uses vowel letters in the course of words, and according to the discretion of readers and the different regions the same word is pronounced with different sounds and accents.” In other words, there was much local variation in the realization of the vowels. One may wish to compare the situation in colloquial Arabic, where slight changes in vowels are noticeable in its various dialects (for example, the definite article can be [al], [el], [il], [əl], or [l]).

Below we present the vowel system according to the Tiberian Masoretic system. We begin with the long vowels, which are far simpler in their historical development, then move to the short vowels, and conclude with a treatment of the diphthongs.

5.6.1. Long vowels

Typically, the Proto-Semitic long vowels retain their basic pronunciation in all environments. Thus, /î/ is always [î], and /û/ is always [û]. The only area of fluctuation is with /â/. When Semitic cognates indicate /â/, the Hebrew reflex typically will be /ô/, though sometimes the /â/ is retained. Thus, for example, Arabic *lâ* = Hebrew *lô* ‘no’; Arabic *salâm* = Hebrew *šālôm* ‘peace’; etc., but Arabic *ṭabbâḥ* = Hebrew *ṭabbâḥ* ‘cook’; etc.

5.6.2. Short vowels

The above discussion (§ 5.6) about the numerous vowel allophones refers most importantly to the short vowels. The Tiberian Masoretic notation system reflects different realizations of the three original vowels /a/, /i/, and /u/, depending on the kind of syllable in which the vowel occurs and depending on the accent.

If the short vowel occurs in an accented syllable, or in an unaccented open syllable immediately preceding the accent, the following developments occur (I include the name of the Hebrew vowel, its Tiberian symbol in parentheses, and the traditional transliteration in italics):

/a/ > [ɔ] *qameṣ* (֫) *ā*

/i/ > [e] *ṣere* (ֿ) *ē*

/u/ > [o] *holem* (׀) *ō*

If the short vowel occurs in an unaccented closed syllable, typically the original pronunciation is not affected, but with two of the vowels there is the possibility of an allophone. Thus:

/a/ > [a] *pataḥ* (ֿ) *a*

/i/ > [i] *ḥiriq* (ֿ) *i*

or

/i/ > [ɛ] *segol* (ֿ) *e*

/u/ > [u] *ṣureq* (ֿ) *u*

or

/u/ > [ɔ] *qameṣ* (֫) *o*

Different environments usually will determine whether /i/ > [ɛ] as opposed to [i], and whether /u/ > [ɔ] as opposed to [u]. For example, if the vowel is followed by a geminated consonant, one can expect /i/ > [i], e.g., *libbî* ‘my heart’, and /u/ > [u], e.g., *kullām* ‘all of them’; by contrast witness /i/ > [ɛ] in *leb-yām* ‘heart of the sea’, and /u/ > [ɔ] in *kol-ʾiṣ* ‘every man’.

If the short vowel occurs in an open syllable more than one syllable before the accent, then the vowel is reduced to shwa [ə] (noted by ְ). If, however, the consonant involved is a pharyngeal or a laryngeal, then an auxiliary

vowel is necessary (often called “compound vowel,” due to its orthographic representation in the Masoretic system) (see § 5.5.3). The auxiliary vowel is halfway between a true shwa and the corresponding short vowel. Thus, using the traditional transliteration of Hebrew grammarians, /a/ > *ā* (אֵ), /i/ > *ē* (יֵ), and /u/ > *ō* (וֵ).

We illustrate this whole process with one example. The word for ‘word’ in Hebrew is [dōvór], with original short vowel /a/ in both syllables. The first [ɔ] occurs because it appears in an unaccented open syllable immediately preceding the accent; the second [ɔ] occurs because it appears in an accented syllable. In the expression ‘the word of Esther’ [dəvar-’estér], the two words together have only the one accent, at the end of the expression. The first /a/ vowel now appears in an unaccented open syllable more than one syllable before the accent, and thus is reduced to shwa. The second /a/ vowel now appears in an unaccented closed syllable and thus is realized as [a].

Note that one Hebrew vowel sign, the *qameṣ* (ִ), is transliterated as *a* when it derives from an /a/ vowel, but is transliterated as *o* when it derives from an /u/ vowel. This reflects the realization of this vowel according to the Jews of most Arab lands and according to standard Israeli pronunciation today. However, the Masoretic notation clearly demonstrates a single pronunciation for this vowel, which most accurately is [ɔ] and which is realized thus by the Jews of Europe and of Yemen. This demonstrates that the short vowel /a/, when it was accented and when it appeared in an open syllable immediately preceding the accent, was raised to a quality approaching the short vowel /o/. Such a process is in fact clearly indicated for Phoenician, and was no doubt true of ancient Hebrew as well, at least in the pronunciation tradition which emerged among the Tiberian Masoretes. It parallels the case of the long vowel /â/ shifting to /ô/; thus we may wish to postulate a general drift in this direction in ancient Hebrew and Phoenician.

It is important to note that the above charting of rules governing the short vowels is not to be taken as hard and fast. As in most languages, also in Hebrew, /a/ is the most stable vowel. When an /i/ vowel or an /u/ vowel is present, often the above rules will be violated. For example, **burâš* > [bərôš] ‘juniper, cypress’ shows reduction of the /u/ vowel to shwa, even though the open syllable in which it occurs immediately precedes the accent. By contrast, of similar nominal pattern is **šalâš* > [šəlôš] ‘three’, with the /a/ vowel retaining its character (actually, with raising to [ɔ], as discussed in the preceding paragraph).

Similarly, auxiliary vowels can arise after consonants which are not pharyngeals or laryngeals. For example, /u/ does not reduce to *shwa* in the word

haggōrānôt ‘the threshing floors’; rather it appears as *ō*. This is due to the circumstance of back vowel /u/ following the velar consonant /g/. Instead of reducing fully to *shwa*, as normally would be expected in the case of an unaccented open syllable more than one syllable before the accent, /u/ retains part of its original quality (i.e., as a back vowel) following a consonant pronounced in the back of the mouth (i.e., the velar /g/).

5.6.3. Diphthongs

Two diphthongs are reconstructed for ancient Hebrew in its earliest stage: [aw] and [ay]. In some cases, for example, in final position, these diphthongs remain unchanged, e.g., *qāw* ‘line’, *ḥay* ‘alive’ (though with the former note again the raising of the vowel to *ā* = [ɔ]). Typically, however, one of two changes occurs. Either an anaptyctic vowel is inserted, thus, e.g., **mawt* > [mōwet] ‘death’ (or [mōwet] showing fricativization), **bayt* > [bayit] ‘house’ (or [bayit] showing fricativization) (again note the raising of the vowel in the former example); or monophthongization occurs.

Monophthongization in Hebrew almost always means [aw] > [o] (traditionally transliterated as *ō*), and [ay] > [e] (traditionally transliterated as *ē*), e.g., **yawm* > [yom] ‘day’, **bayda* > [beṣo] ‘egg’. However, in a small number of instances, these two diphthongs monophthongize to [ɔ] (traditionally transliterated as *ā*). Examples of this latter process may be localized to two geographical regions in Israel: the northern part of the country (Galilee) and a small pocket in southern Judah (northern Negev).

5.6.4. Vowel letters

While a treatment of the vowel letters more properly belongs to a discussion of orthography rather than of phonology, a brief mention of them is appropriate. First, however, a basic overview of the problem is necessary. The oldest Hebrew inscriptions do not indicate the vowels; instead the 22-letter alphabet represents only the consonants. From the 8th century B.C.E. onward, according to the standard view, the practice arose to utilize certain letters, namely, <h>, <w>, and <y>, to indicate vowels (first only final vowels were indicated, later the practice was extended to mark medial vowels as well). When used in this manner, these letters (as already has been mentioned, see § 5.3) are known as *matres lectionis* or “vowel letters.” By the 1st century B.C.E., this practice had increased so greatly, that in some of the Dead Sea Scrolls from this period virtually all vowels are marked by the aforementioned letters.

The Masoretic text presents a middle ground. Even though our earliest Masoretic manuscripts are from the early Middle Ages, they must go back to much older prototypes, because generally they are much more conservative in their use of the vowel letters than are the Dead Sea Scrolls of a millennium earlier. Two examples will suffice: in the Bible [lo'] 'no, not' is spelled regularly <l'> and more rarely <lw'>; in the Dead Sea Scrolls there are about 400 cases of <lw'> and about 100 cases of <l'>. Similarly, in the Bible [kol] ~ [kɔl] is spelled regularly <kl> and in only one case <kwl>; in the Dead Sea Scrolls there are about 700 cases of <kwl> and only about three dozen cases of <kl>.

Most scholars have concluded that the use of the vowel letters in the Masoretic text is arbitrary, i.e., they have no phonetic significance. According to this theory, whether a given word is spelled with vowel letter or without indicates nothing about the pronunciation of the word. However, close analysis often reveals a remarkable degree of consistency in spelling variation, and this consistency, it has been argued, indicates that the vowel letters indeed do tell us something about the actual pronunciation of the Hebrew word. According to this view, the vowel letters <w> and <y> indicate an off-glide. For example, <qwl> 'voice' would have been pronounced [qo^wl], with the allophonic off-glide, but <hql> 'the voice' would have been pronounced [haqqol]. The majority view has so dominated the field of Hebrew linguistics that little regard has been paid to the minority view. Further research on this issue remains a desideratum, but an open mind should be kept once the idea of allophonic off-glides is countenanced.

5.7. Historical changes concerning the vowels

5.7.1. /i/ > /a/ in an originally closed accented syllable

This law is known as Philippi's Law. An original /i/ vowel shifts to /a/ in an originally closed accented syllable (that is, a syllable that was closed even in its proto-form [as opposed to a closed syllable brought about by some other historical development]). Thus, for example, Proto-Semitic **gint* > **gitt* (via assimilation, see § 5.8.2) > **git* (with surrendering of word-final gemination) > [gat] 'winepress, olivepress'. In Akkadian transcriptions of the city in Canaan by this name, dating to as late as ca. 720 B.C.E., the form is still *Gint* (or *Gimt* [with partial dissimilation]). In the Septuagint of ca. 200 B.C.E., the rendering reflects [get], and in the Masoretic text the pronunciation is [gat]. Accordingly, we are able to trace the historical development of this shift,

though the Septuagint rendering is too equivocal ([gɛt] apparently halfway between earlier [git] and later [gat]) to allow us to pinpoint the century in which Philippi's Law occurred.

5.7.2. /a/ > /i/ in an originally closed unaccented syllable

This law does not have an official name, but it may be called the corollary to Philippi's Law. An original /a/ vowel shifts to /i/ in an originally closed unaccented syllable (again, that is, a syllable that was closed even in its proto-form [as opposed to a closed syllable due to some other historical development]). Thus, for example, **magdal* > [migdal] 'tower' (also a toponym 'Migdal'); **šamšōn* > [šimšōn] 'Samson'; etc. In the Septuagint and the New Testament (1st century C.E.), the Greek renderings of proper names reflect the original /a/ vowel (witness our English Samson, Mary Magdalene, etc.). Jerome (ca. 400 C.E.) still has *Magdal* in his Latin translation of the Bible. The Masoretic text reflects the shift to /i/ at some point within the following four and a half centuries. Thus, we may date this shift to sometime between 400 C.E. and 850 C.E.

5.8. Varia

5.8.1. Metathesis

The most consistent case of metathesis occurs in the Hitpa'el form of the verb, when the first root consonant is any of the sibilants, /s/, /z/, /ʃ/, /ʒ/, or the lateral fricative /ɬ/. In such cases, the /t/, which forms part of the morphology of this verbal stem and which normally precedes the first root consonant, interchanges with the above consonants. For example, **ʔetšammer* > [ʔɛštammer] 'I guard myself'.

Other examples of metathesis are the word pairs [kɛvɛʃ] ~ [kɛʃɛv] 'sheep', and [šimlɔ] ~ [šalmɔ] 'article of clothing', both of which interestingly contain the lateral fricative /ɬ/.

5.8.2. Assimilation

Regressive assimilation occurs with vowelless /n/, except before pharyngeals and laryngeals. Thus, for example, to use an item noted earlier, **gint* > **gitt* (eventually shifting to [gat]) 'winepress, olivepress'. Similarly, **yandur* eventually emerges as [yiddor] 'he vows'. Note also the same phenomenon with vowelless /l/ in various forms of the verb *lqh* 'take' (e.g., **yilqah* > *yiqqah* 'he takes'); and with vowelless /d/ preceding its voiceless counterpart /t/. A

regular example of the latter is **ʾahadt* > [ʾaḥat] ‘one’ (fem.). Another unique example occurs in **lalidt* > **laladt* (via Philippi’s Law) > **lalatt* > [lalat] (with surrendering of final gemination) ‘to give birth’, a form which occurs only once in the Bible (the normal form is [lələdet], or with fricativization [lələdɛt], arrived at through different means).

Partial progressive assimilation occurs in the Hitpa‘el form of the verb, when the first root consonant is /z/ or /ṣ/ and it precedes /t/ (see also § 5.8.1). No examples with /z/ occur in the Bible, but from post-biblical Hebrew we may cite **hiztayyef* > [hizdayyef] ‘be forged’, in which /t/ shifts to /d/ because of the preceding /z/. One example with /ṣ/ occurs in the Bible: **niṣṭaddaq* > [niṣṭaddaq] ‘(how) shall we justify ourselves’, in which /t/ shifts to /ṭ/ because of the preceding /ṣ/.

5.8.3. *Prothetic vowel*

The pronunciation of initial consonant clusters is assisted by the placement of a prothetic vowel. The best example is the attestation of both [zəroaʿ] and [ʕzroaʿ] ‘arm’, though the latter may be limited to specific regional dialects. Another example is [ʕəṣbaʿ] ‘finger’, which from the cognate evidence (especially Egyptian *ḏb*) can be shown to be originally without the initial [ʕ-].

5.8.4. *Anaptyxis*

The presence of anaptyctic vowels has been noted on several occasions above (see §§ 5.5.3, 5.6.3). One further example occurs in the creation of the “segolate” nouns, e.g., **dalt* > *dalet* (attested in Hebrew in sentence positions requiring a pause, e.g., at the end of a verse) > *delet* [dɛlet] (with vowel harmony) ‘door’. Greek and Latin transliterations of such words tend to show the forms without anaptyxis, though they do so inconsistently. In any case, this development most likely occurred in the 1st millennium C.E.

5.8.5. *Stress*

Stress in Hebrew at times is phonemic, e.g., [rəḥel bəʔɔ] ‘Rachel is coming’ vs. [rəḥel bəʔɔ] ‘Rachel came’.

Bibliography

Blau, Joshua. 1970. *On Pseudo-Corrections in Some Semitic Languages*. Jerusalem: Israel Academy of Sciences and Humanities.

- . 1982. *On Polyphony in Biblical Hebrew*. Proceedings of the Israel Academy of Sciences and Humanities 6/2. Jerusalem: Israel Academy of Sciences and Humanities.
- Garr, W. Randall. 1990. "Interpreting Orthography." In W. H. Propp, B. Halpern, and D. N. Freedman, eds., *The Hebrew Bible and Its Interpreters*, pp. 53–80. Winona Lake, Ind.: Eisenbrauns.
- Kutscher, E. Y. 1982. *A History of the Hebrew Language*, ed. Raphael Kutscher. Jerusalem: Magnes; Leiden: Brill.
- Levin, Saul. 1988. "The Hebrew of the Pentateuch." In Y. L. Arbeitman, ed., *Fucus: A Semitic/Afrasian Gathering in Remembrance of Albert Ehrman*, pp. 291–323. Amsterdam: Benjamins.
- Moscatti, Sabatino, et al. 1964. *An Introduction to the Comparative Grammar of the Semitic Languages*. Wiesbaden: Harrassowitz.
- Qimron, Elisha. 1986. *The Hebrew of the Dead Sea Scrolls*. Atlanta: Scholars Press.
- Rendsburg, Gary. A. 1990a. *Linguistic Evidence for the Northern Origin of Selected Psalms*. Atlanta: Scholars Press.
- . 1990b. "Monophthongization of *aw/ay* > *ā* in Eblaite and in Northwest Semitic." *Eblaitica* 2: 91–126.
- Segal, M. H. 1927. *A Grammar of Mishnaic Hebrew*. Oxford: Clarendon.
- Steiner, Richard. C. 1977. *The Case for Fricative-Laterals in Proto-Semitic*. New Haven: American Oriental Society.
- . 1982. *Affricated Šade in the Semitic Languages*. New York: American Academy for Jewish Research.
- Waldman, Nahum M. 1989. *The Recent Study of Hebrew*. Cincinnati: Hebrew Union College Press; Winona Lake, Ind.: Eisenbrauns.
- Zevit, Ziony. 1980. *Matres Lectionis in Ancient Hebrew Epigraphs*. Cambridge, Mass.: American Schools of Oriental Research.

This page intentionally left blank.

Tiberian Hebrew Phonology

Geoffrey Khan

University of Cambridge

Until the second half of the first millennium A.D., the text of the Hebrew Bible was transmitted in a form of writing that represented the consonantal phonemes but left the majority of the vowels and also consonantal gemination without graphic expression. When the Bible was read aloud, the reader followed a tradition of pronunciation that was transmitted orally and changed with the passage of time. At some period between the seventh and ninth centuries A.D., a circle of scholars in Tiberias known as Masoretes recorded in written form many of the missing details of the pronunciation of Biblical Hebrew, including the vowels, consonantal gemination, and even the distinction between the allophones of some of the consonantal phonemes. They also recorded the musical cantillation of the reading tradition. The system of signs created by the Tiberian Masoretes to represent these details is known as the Tiberian vocalization system. During the Middle Ages other vocalization systems were developed, which used different signs. The Tiberian system, however, became standardized and gradually replaced the others.

We must distinguish the Tiberian vocalization system from the original Tiberian Hebrew pronunciation, which it was designed to represent. This was the pronunciation of Hebrew which was used in the traditional reading of the Bible in the region of Tiberias during the seventh–ninth centuries A.D. Whereas the Tiberian vocalization tradition has survived in written form, the Tiberian pronunciation of Hebrew, which was an orally transmitted tradition, is extinct. None of the pronunciation traditions of the Hebrew Bible that are in use among Jewish communities today derive from the Tiberian pronunciation.

The original Tiberian pronunciation that lies behind the vocalization signs can be reconstructed from several sources. These include:

1. Masoretic and grammatical texts. Of primary importance are the texts from Palestine, especially the work *Hidāyat al-qāri* ‘Guide for the reader’. The grammarians from medieval Spain sometimes describe the articulation

of a sound in greater detail than the Eastern sources. Their descriptions have to be treated with caution, however, since they could in some cases reflect a local type of pronunciation that differed from the Tiberian.

2. Transcriptions of the Tiberian pronunciation tradition into Arabic script which are found in medieval manuscripts written by Karaites (a medieval sect of Judaism).

3. The use of Hebrew letters and Tiberian vocalization signs to represent other languages. Of particular importance are medieval texts that represent Arabic in this way.

In this chapter an attempt is made to present the main features of the Tiberian pronunciation tradition based on the latest research on the aforementioned medieval sources.

6.1. Consonants

The letters are discussed in alphabetical order.

ʾAleṣ (א). /ʔ/

Phonetic realization: Glottal plosive [ʔ].

Beṯ /b/

Phonetic realization: Two allophones: (1) (ב) Voiced bilabial stop [b] and (2) (ב) voiced labiodental [v].

Hidāyat al-qāri describes the [b] allophone as primary (ʾaṣl) and the [v] allophone as secondary (*farʿ*) (fols. 8b, 10a; cf. Eldar 1980–81: 254 n. 58).

Gimel /g/

Phonetic realization: Two allophones: (1) (ג) Voiced velar stop [g] and (2) (ג) voiced uvular fricative [ɣ].

Hidāyat al-qāri describes the [g] allophone as primary (ʾaṣl) and the [ɣ] allophone as secondary (*farʿ*) (fol. 8b; cf. Eldar 1980–81: 254 n. 58).

Daleṯ /d/

Phonetic realization: Two allophones: (1) (ד) Voiced post-dental stop [d] and (2) (ד) voiced post-dental fricative [ð].

The *Hidāyat al-qāri* describes the [d] allophone as primary (ʾaṣl) and the [ð] allophone as secondary (*farʿ*) (fol. 8b; cf. Eldar 1980–81: 254 n. 58).

The medieval scholar Isaac Israeli (9th–10th centuries A.D.), who had expert knowledge of the Tiberian reading tradition, is said to have pronounced [ð] with a secondary “emphatic” (i.e. velarized or uvularized) articulation [ð̤] in two words, viz. [ʔappað̤no:] ‘his palace’ (Dan. 11:45) and [va:jjəð̤raʿxu:] ‘and they have bent’ (Jer. 9:2) (cf. Schreiner 1886: 221; Mann 1931–35, 1: 670 n. 106; Dukes 1845–46: 9, 73; Grossberg 1902: 24).

He (ה) /h/

Phonetic realization: Glottal fricative [h].

Waw (ו) /w/

Phonetic realization: Two allophones: (1) Labiodental [v] and (2) labio-velar semivowel [w].

The usual realization of /w/ was [v]. The allophone [w] occurred when the letter was preceded or followed by a *u* vowel, e.g. [ufuw'wɔ:] 'and Puwwa' (proper name) (Gen. 46:13), [vajjiʃta:ħa'wu:] 'and they prostrated themselves' (Deut. 29:25), [tɔ:'wu:] 'they span' (Ex. 35:26) (see David ben Abraham al-Fāsī 1936, 1:451–52; Mishael ben Uzziel 1965: 20; Eldar 1978, 1:85, 1980–81: 259, 1984: 10–11).

Zayin (ז) /z/

Phonetic realization: Voiced alveolar sibilant [z].

The *Hidāyat al-qāri* mentions a variant of the letter *zayin* which is referred to by the Tiberian scholars as *zāy*¹ *makrūk* (Eldar 1984–85: 32).² The epithet *makrūk* was used by the Tiberian scholars to describe also a variant type of *reš*. It apparently referred to an emphatic (i.e. velarized or uvularized) articulation of the letter (cf. Khan to appear a). It appears, therefore, that *zayin* had an emphatic allophone [z̤], though its distribution is unknown.

Ḥeṭ (ח) /ħ/

Phonetic realization: Unvoiced pharyngeal fricative [ħ].

Ṭeṭ (ט) /t/

Phonetic realization: emphatic (i.e. velarized or uvularized) unvoiced alveolar plosive [t̤].

Yod (י) /j/

Phonetic realization: palatal unrounded semivowel [j].

According to one medieval source (Saadya Gaon 1891: 42–43), the Tiberians pronounced geminated *yod* like Arabic *jīm*, i.e. as a voiced palatal stop [j̤] (cf. Roman 1983: 101–6, 218), which had the same place of articulation as *yod* [j]. This was the result of strengthening the articulation of [j] to a stop.

Kaṭ (כ) /k/

Phonetic realization: Two allophones: (1) (כּ) Unvoiced velar stop [k] and (2) (כּ) unvoiced uvular fricative [χ].

1. The *Hidāya* uses the Arabic letter name.

2. The Yemenite orthoepic treatise known as the Hebrew *Maḥberet ha-Tījān*, which was based on the long version of the *Hidāya*, contains a similar statement (1870: 81, cf. Morag 1959–60: 219 n. 45): *wkn yš lhm zyn nqr' mkrwk v'ynw ydw' šlynw* 'They (i.e. the Jews of Palestine) have a *zayin* called *makrūk*, but it is unfamiliar to us (i.e. the Jews of Yemen)'.

We know from Greek transcriptions that in the first half of the first millennium A.D., plosive *kaḇ* was pronounced with aspiration [k^h] (cf. Kutscher 1965: 24–35). This was likely to be the case also in the Tiberian pronunciation tradition.

Hidāyat al-qāri describes the [k] allophone as primary (ʿaṣl) and the [χ] allophone as secondary (*farʿ*) (fols. 8b, 10a; cf. Eldar 1980–81: 254 n. 58).

Lamed (ل) /l/

Phonetic realization: Voiced alveolar lateral continuant [l].

Mem (م) /m/

Phonetic realization: Voiced bilabial nasal [m].

Nun (ن) /n/

Phonetic realization: Voiced alveolar nasal [n].

Samek (ס) /s/

Phonetic realization: Unvoiced alveolar sibilant [s].

ʿAyin (ע) /ʕ/

Phonetic realization: Voiced pharyngeal fricative [ʕ].

Pe /p/

Phonetic realization: Two allophones: (1) (פ) Unvoiced bilabial stop [p] and (2) (פּ) unvoiced labiodental fricative [f].

The *Hidāyat al-qāri* describes the [p] allophone as primary (ʿaṣl) and the [f] allophone as secondary (*farʿ*) (fol. 8b; cf. Eldar 1980–81: 254 n. 58).

Saadya refers to the existence of a “hard” *pe* in the word [ʔappaðʕno:] ‘his palace’ (Daniel 11:45), which he describes as between plosive *beṭ* and plosive *pe* (1891: 42). This appears to be referring to an unaspirated, fortis realization of [p]. One may infer from this that the voiced stop *beṭ* was unaspirated whereas the normal unvoiced stop *pe* was aspirated. We know from Greek transcriptions that in the first half of the first millennium A.D. plosive *pe* was pronounced with aspiration (cf. Kutscher 1965: 24–35). This appears also to have been the case in the Tiberian pronunciation tradition. Dunash ben Tamim reports that the scholar Isaac Israeli (9th–10th centuries), who was “an expert in the reading of the Tiberians,” pronounced the *daleṭ* in the word [ʔappaðʕno:] as the emphatic (velarized or uvularized) Arabic letter *zāʾ* (cf. Schreiner 1886: 221; Mann 1972: 670 n. 106; Dukes 1845–46: 9, 73; Grossberg 1902: 24). This implies that the ‘hard’ *pe* was also emphatic [p] (cf. Steiner: 1993: 551–61).

Ṣade (צ) /s/

Phonetic realization: Unvoiced emphatic (velarized or uvularized) alveolar sibilant [s̠].

Ibn ǰaldūn refers to a voiced allophone of *šade* [z] in the pronunciation of the name **יְהוֹשֻׁעַ**, i.e. [ʔamazʔjo:hu:] (cf. Schreiner 1886: 254).

Quṣ (ק) /q/

Phonetic realization: Unvoiced uvular plosive [q].

According to the *Hidāyat al-qārī*, it was articulated with the middle of the tongue (*waṣṭ al-lisān*), i.e. somewhere on the tongue between the back third and the tip (fols. 10a–10b, ed. Eldar 1980–81, lines 61–72). This was further forward than fricative *kaṣ* and *gimel*, which were articulated with the back third of the tongue.

Reš (ר) /r/

Phonetic realization: Two allophones: (1) Voiced uvular roll [R] or frictionless continuant [ʀ] and (2) emphatic (velarized or uvularized) linguo-alveolar roll [r].

The basic articulation of the Tiberian *reš* was an uvular roll [R] or frictionless continuant [ʀ]. It was realized as the allophone [r] in the environment of the alveolar consonants /d/, /z/, /s/, /t/, /t̪/, /s/, /l/, /n/. The precise rule given in the medieval sources is that [r] occurred when preceded by /d/, /z/, /s/, /t/, /t̪/, /s/, /l/, /n/ or followed by /l/, /n/ and when the *reš* was either in contact with these letters or at least in the same syllable, e.g. [daʔ'ko:] 'his way' (Gen. 24:21), [jizʔo:q] 'he scatters' (Isa. 28:25), [liʔho:s] 'to wash' (Gen. 24:32), [saʔ'ne:] 'rulers' (Josh. 13:3). This allophone of *reš* was alveolar, by assimilation to the adjacent alveolar consonants. It can be established from medieval sources that it was also emphatic (velarized or uvularized) (see Khan to appear a).

Sin (ש) /s/

Phonetic realization: Unvoiced alveolar sibilant [s].

This had the same pronunciation as *sameḵ* in the Tiberian pronunciation tradition.

At an earlier historical period, *sin* and *sameḵ* were distinct in pronunciation and represented separate phonemes, as is shown by minimal pairs such as **סַר** [sa:ʔ] 'stubborn' and **שַׁר** [sa:ʔ] 'rule, captain'.

Šin (שׁ) /ʃ/

Phonetic realization: Unvoiced palato-alveolar fricative [ʃ].

Taw /t/

Phonetic realization: Two allophones: (1) (ת) Unvoiced alveolar stop [t] and (2) (ת) unvoiced alveolar fricative [θ].

The *Hidāyat al-qārī* describes the [t] allophone as primary (ʾašl) and the [θ] allophone as secondary (*fārʿ*) (fol. 8b; cf. Eldar 1980–81: 254 n. 58).

We know from Greek transcriptions that in the first half of the first millennium A.D. plosive *taw* was pronounced with aspiration (cf. Kutscher 1965: 24–35). This was likely to be the case also in the Tiberian pronunciation tradition.

6.1.1. *Distribution of the allophones of /b/, /g/, /d/, /k/, /p/, /t/*

In general the fricative allophones of these letters (i.e. [v], [ɣ], [ð], [χ], [f], and [θ] respectively) occurred after a vowel when the letter was not geminated, e.g. [ra:v] ‘much’, [ʃɔ:vɑ:ɾ] ‘he broke’, [jiʃka'vu:] ‘they lie down’. In many cases, however, the preceding vowel had been elided some time in the history of the language before the period of the Masoretes but the letter nevertheless remained a fricative, e.g. [baʃɔχ'vo:] < **bašuku'bō* ‘when he lies down’, [mal'χe:] < **mala'kē* ‘kings’, [ʃɔ:χ'vu:] < **šāka'bū* ‘they lay down’. In a few such cases a plosive and a fricative are in free variation, e.g. [riʃ'fe:] and [riʃ'pe:] ‘flames’. The distribution of the plosive and fricative allophones, therefore, is not completely predictable from the phonetic context in Tiberian Hebrew, since it is an alternation that was inherited from an earlier stage of the language.³

In theory the phonetic processes described above could have given rise to a phonemic opposition between the plosive and fricative forms of the letters. However, no certain minimal pair that proves this opposition is attested in the corpus of the Hebrew Bible. Z. Harris (1941: 143–67) proposed the hypothetical minimal pair [ʔal'fe:] ‘thousands’ vs. [ʔal'pe:] ‘two thousand’. The form of the second word in the pair is deduced from what we know about Hebrew morphology but is not attested.

6.1.2. *Consonant gemination*

This is marked in the Hebrew script by placing a dot in the letter known as *dageš*. According to the *Hidāyat al-qāri*, “*dageš* makes the letter heavy.” This “heaviness” of letters is brought about by increased muscular pressure of speech organs (*Hidāyat al-qāri* fol. 9a–9b, ed. Eldar 1980–81, lines 15–16, 37–38). A geminated consonant, therefore, was pronounced with greater pressure than its ungeminated counterpart. Some consonants could not be geminated. These included the laryngeals (/ʔ/, /h/) and pharyngeals (/ʕ/, /ħ/) and also /r/, except in a few isolated cases.

3. This is a simplified account of the distribution of the allophones of /b/, /g/, /d/, /k/, /p/, /t/. For a more detailed description see Yeivin 1980: 285–96.

6.2. Vowels

Tiberian Hebrew had the vowel system shown in Table 6-1.

Table 6-1. Tiberian Hebrew Vowels

i	u
e	o
ε	ɔ
a	

Pataḥ (ֿ) /a/

Phonetic realization: Open, unrounded. There was no phonemic opposition between front and back vowels in the open position, so the allophonic scatter of /a/ is likely to have included both front [a] and back [ɑ] qualities. Evidence for this can be found in Judaeo-Arabic texts with Tiberian vocalization.⁴

Segol (ֿ) /ε/

Phonetic realization: front, half-open unrounded [ε].

Qameṣ (ֿ) /ɔ/

Phonetic realization: back, half-open rounded [ɔ].

Šere (ֿ) /e/

Phonetic realization: front, half-close unrounded [e].

Ḥolem (ֿ) /o/

Phonetic realization: back, half-close rounded [o].

Ḥireq (ֿ) /i/

Phonetic realization: front, close, unrounded [i].

Šureq (ֿ), qibbuṣ (ֿ)⁵ /u/

Phonetic realization: back, close, rounded [u].

6.2.1. Vowel length

Vowel length is in most cases predictable from syllable structure and the placement of stress. Meaningful contrasts between words were not usually

4. In one text (T-S Ar. 8.3), for instance, which uses both *pataḥ* and *qameṣ* signs, *pataḥ* is used to represent Arabic *fatḥa* both in the environment of emphatic consonants, where it would be expected to have had a back quality [ɑ] (e.g. [ʔaʕʕam]), and also in the environment of non-emphatics, where a front quality [a] would have been expected (e.g. [watafi]). The *qameṣ* sign is used in this text to represent a back vowel somewhere in the region of mid vowels [ɔ] and [o] which resulted from the contraction of the diphthong [aw], e.g. [foːq].

5. These are orthographic variants of the same vowel.

made by differences in vowel length alone. Differences in length are in virtually all cases relatable to differences in syllable structure or stress placement. Length was not an independent contrastive feature of vowels. The vowel *qameṣ* may have been an exception, since pairs of words can be found in which a contrast of meaning appears to have been made only by a difference in length of the vowel, e.g. [ʔɔχ'ɫɔ:] 'food' vs. [ʔɔ:χɫɔ:] 'she ate'. Possible other minimal pairs are words such as [dɔ'mi:] 'silence' and [dɔ:'mi:] 'my blood'. The validity of both such minimal pairs, however, is not completely certain (see below).

The basic contexts for the occurrence of a long vowel are (1) a stressed syllable or (2) an open unstressed syllable. Examples: ['mæ:lɛχ] 'king', [ji'ma:f] 'he hears', [ha:'hu:] 'that'. Many words carry a secondary stress in addition to the main stress, e.g. [hɔ:ʔɔ:'ðɔ:m] 'the man', [ni:θhakkamɔ:] 'let us deal wisely' (Ex. 1:10).

As has been remarked, a vowel in an unstressed closed syllable was, on principle, short. If, however, it was followed by a series of contiguous consonants of relatively weak articulation (e.g. /ʔ/, /h/, /ʕ/, /ħ/, /j/, /n/, /l/), then the vowel was sometimes lengthened even when not stressed. This occurred in certain prefixes of the verbs [hɔ:'jɔ:] 'he was' and [hɔ:jɔ:] 'he lived', namely the [i] of prefixes before [h]/[ħ], e.g. [ji:hjɛ:] 'he will be', and the [a] of the conjunctive prefix [va] before [j], e.g. [va:jhi:] 'and it was'. It is occasionally found elsewhere, e.g. [ha'fɔ:ma:f 'fɔ:m] 'did any people hear?' (Deut. 4:33).

The duration of long vowels varied considerably. From the medieval sources we are able to infer the existence of several different degrees in the relative duration of long vowels. Most of these were conditioned by differences in stress, vowel height, or consonantal strength. We shall mention here some of the conditions of these variations that are known in the present state of research.⁶ This list does not include all the variations that we have evidence for. There were likely to have been, moreover, a number of other variations for which we have no evidence from the extant sources.

1. Stressed long vowels were longer than unstressed long vowels, e.g. in the word [ha:'hu:] 'that' the [u:] was longer than the [a:].

2. A long vowel with secondary stress was longer than a long vowel in an unstressed syllable, e.g. in the word [hɔ:ʔɔ:'ðɔ:m] 'the man' the second [ɔ:] was shorter than the other two.

3. A close vowel [i, u] in a closed syllable with secondary stress was shorter than an open vowel [a] in the same conditions, e.g. in the words [ni:θhakk-

6. For the evidence for these variations see Khan 1987, 1989, 1994b.

ka'mo:] 'let us deal wisely' (Ex. 1:10) and [va:ttiṣpa'ne:hu:] 'and she hid him' (Ex. 2:2), the [i:] vowel of the first was shorter than the [a:] of the second.

4. The close vowel [i] of prefixes of the verbs [hɔ:jɔ:] 'he was' and [hɔ:jɔ:] 'he lived' was shorter than the open vowel [a] in prefixes of these verbs, e.g. in the words [ji:hje:] 'he will be' and [va:jhi:] 'and he was' the [i:] of the first was shorter than the [a:] of the second.

5. The close vowel [i:] of the prefixes of the verbs [hɔ:jɔ:] 'he was' and [hɔ:jɔ:] 'he lived' was shorter than [i:] in a stressed syllable or an unstressed open syllable but longer than [i:] in a closed syllable with secondary stress, e.g. in the words ['ʔi:m] 'if', [ji:hje:] 'he will be', and [ni:θhakka'mo:] 'let us deal wisely' the three [i:] vowels were of decreasing degrees of length.

6. The [a:] vowel in prefixes of the verbs [hɔ:jɔ:] 'he was' and [hɔ:jɔ:] 'he lived' (e.g. [va:jhi:] 'and he was') and in other words before two weak consonants (e.g. [ha'ʔo:ma:f ʔo:m] 'did any people hear?' Deut. 4:33) was longer than an [a:] vowel in a closed syllable with secondary stress (e.g. in [va:ttiṣpa'ne:hu:] 'and she hid him' Ex. 2:2).

6.2.2. Syllable structure and the *šewa*

In the Tiberian pronunciation tradition, many short vowels occurred in open syllables, e.g. [jiʃma'ru:] 'they guard', [ja:ʔa'se:] 'he does'. These were represented in the vocalization system by the *šewa* sign or one of the *ḥaṭeṭ* signs. These were different from the regular vowel signs. From the Masoretic sources and Judaeo-Arabic texts with Tiberian vocalization, we know that these vowels were equivalent in length to short vowels in unstressed closed syllables (see Khan 1987: 37–39, 1992: 105–11). Does the occurrence of these short vowels in apparently open syllables contradict the vowel length principle stated above?

According to the medieval Masoretic sources, a consonant with one of these vowels did not constitute a syllable. In a word such as [tiṣpa'ru:] 'you count', the syllable structure would be, according to the medieval sources, [tis-pa'ru:]. This concept of the syllable reflects the phonotactic rules of Tiberian Hebrew and corresponds to the phonotactic definition of syllables espoused in modern times by linguists such as Pulgram (1970: 40ff.). The basic principle of Pulgram's definition is that a sequence of consonant and vowel segments has the status of a syllable only if the onset of the sequence can stand in word-initial position and the coda (i.e. closure) can stand in word-final position. There is no structural reason why it cannot stand by itself as a word. In the medieval Tiberian reading tradition of Biblical

Hebrew, a short vowel did not occur in word-final position. According to this definition, therefore, the sequence consonant + short (CV) vowel did not have the status of a syllable. Only consonants and long vowels could occur in word-final position, and so only these could constitute permissible codas of syllables.⁷ The sequence CV occurred in word-initial position. It could, therefore, form the onset of a syllable. This allowed it to be attached to the beginning of a sequence which had a permissible coda and so had the status of a syllable, viz. CV+CVC or CV+C $\bar{V}$. The sequences CVCVC and CVCV $\bar{V}$, therefore, were regarded by the Masoretes as single syllables.

Rather than denying the status of syllable completely to a CV sequence on the basis of this phonotactic definition, it is helpful to distinguish between principal and dependent syllables. Principal syllables are those that can stand independently, since they have onsets and codas that can open or close an independent word. A dependent syllable is one that cannot stand independently, but only in combination with a following principal syllable. The aforementioned distribution of vowel length, therefore, refers to principal syllables. Any open syllable with a short vowel must be a dependent syllable. This is a phonotactic distinction. It is not usually taken account of by the accent system of Tiberian Hebrew, which counts beats on syllable nuclei between accents without distinguishing between dependent and principal syllables.

The reality of the phonotactic distinction between dependent and principal syllables is reflected by the concept of the syllable that is expressed in the medieval Masoretic literature. It is also reflected by the vocalization system, which represents the vowel nuclei of dependent syllables with signs (*šewa* and *ḥatepim*) that are different from those representing the nuclei of principal syllables. Furthermore, some features of Tiberian Hebrew phonology are sensitive to the distinction. The occurrence pattern of the allophones of Tiberian /r/ is a clear example of this. The apico-alveolar allophone of /r/, i.e. [r], occurred when it was preceded by one of the dental/alveolar consonants /d/, /z/, /t/, /s/, /ʃ/, /t/, /l/, /n/ and when either (a) the *reš* was in direct contact with one of these letters or (b) the *reš* occurred together with one of them in the same syllable, e.g. [daʁkamo:'ni:m] 'drachmas', [vɔ:'ʔez'ʔe:m] 'and I winnowed them' (Jer. 15:7) [bamiz'ʔe:] 'with a pitchfork' (Jer. 15:7), [ʃaʁu:'fɔ:] 'smelted',

7. The only possible exceptions are words ending in a consonantal cluster such as וַיִּשְׁקֵי 'and he watered', וַיִּנְד 'nard'. Some medieval sources state that the second *šewa* in these words was vocalic (e.g. David Qimḥi 1952: 16–17). Most sources, however, state that both *šewas* were silent (e.g. Ibn Janāḥ 1880: 275, Abraham ibn Ezra 1791: 3).

[lim'ta:r] 'through the rain'. When the dental/alveolar was followed by a full vowel the /r/ was realized with the uvular allophone [ʀ], e.g. in [tɔ:'ru:ʂ] 'you run'. How did words such as [lim'ta:r] and [ʂaʀu:'fɔ:] differ from [tɔ:'ru:ʂ]? The most obvious answer is that in [lim'ta:r] and [ʂaʀu:'fɔ:] the *reš* was in the same syllable as the dental/alveolar, whereas in [tɔ:'ru:ʂ] it was in a different syllable.

We may, therefore, elaborate the description of the contexts for the occurrence of a long vowel as follows: A vowel is long if it occurs in a stressed syllable or in an open principal syllable.

There are no phonological oppositions between the vowel of a dependent open syllable CV (represented by vocalic *šewa* or a *ḥatef* sign) on the one hand and zero (represented by silent *šewa*) on the other. The vowel in the syllable CV, therefore, can be regarded as an allophone of zero. It is no doubt for this reason that the Masoretes did not consider vocalic *šewa* to be a vowel and represented it with the same sign as they represented zero. A word such as [ja'vu:] 'sit! (pl.)', therefore, should be represented phonologically as /ʂbu/. There are phonological oppositions, on the other hand, between the vowel of the dependent syllable CV and that of the principal syllable CV, e.g. [ja'vu:] 'sit!' (imperative pl.) vs. [ʃɔ:'vu:] 'they captured'.

In the Tiberian reading tradition, a short vowel in the dependent syllable CV, which was represented by the *šewa* sign, was usually pronounced with the quality of [a]. Where, however, *šewa* preceded a guttural consonant it took the quality of the vowel after the guttural and where it preceded [j] it had the quality of a short [i], e.g. בְּאֵר [be'ɛ:r] 'well', מְאֹד [mo'ɔ:ð] 'very', בְּיוֹם [bi'jo:m] 'on the day' (Baer and Strack 1879: 12–15; Yeivin 1980: 281–82). In places the Masoretes considered that the reader may be uncertain whether to pronounce the *šewa* as vocalic or silent and may have been unsure about the pronunciation of *šewa* where its quality differed from the norm. In such circumstances, the Masoretes added a vowel sign to the *šewa* sign creating a composite sign known as a *ḥatef* sign. The marking of the *ḥatef* signs under the gutturals was fixed in the Tiberian Masoretic tradition, and the Tiberian model codices do not exhibit significant differences. The marking of these signs under the non-gutturals, however, was not fixed, and considerable differences are found in the manuscripts.

Some scholars have claimed that the quality of the *ḥatef* vowels was phonemic on the basis of pairs such as אֲנִיָּה [ʔaniɲ'jɔ:] 'mourning' vs. אָנִיָּה [ʔɔniɲ'jɔ:] 'ship'; חֲלִי [ħa'li:] 'ornament' vs. חָלִי [ħɔ'li:] 'illness'; עָלִי [ʕa'li:] 'go up!' (imperative fem.sg.) vs. עֲלִי [ʕɛ'li:] 'pestle' (cf. Cantineau 1950: 114–16, Garbell 1958–59: 154). If this is the case, they could not be interpreted as

allophones of zero. It will be shown below, however, that the validity of these minimal pairs is doubtful.

Although vowel length is in general predictable from the syllabic context, it would appear that the syllable structure was determined by the length of the vowels. This is because a sequence containing vowels of unspecified length could have been syllabified in various ways, e.g. *tisparu* ‘you count’ could be [tis-pa:-ru:] or [tis-paru:]. The correct syllabification [tis-paru:] could only have been achieved if the length of the vowels had already been fixed.

The length of vowels in the Tiberian pronunciation tradition was determined by the earlier history of the language or by phonetic processes that were operative during the masoretic period. Some long vowels were originally long, e.g. [ko:'he:n] ‘priest’ < **kāhin*. Others were lengthened through phonetic processes that took place at various periods, e.g. lengthening of a vowel in an open syllable before the stress (pretonic lengthening), e.g. [jɔ:'qu:m] ‘he rises’ < **ya'qūm*; the lengthening of stressed vowels, e.g. [mið'bɔ:ɾ] ‘desert’ < **mid'bar*; lengthening of vowel as compensation for the loss or absence of gemination in the following consonant, e.g. [javɔ:'re:χ] ‘he blesses’ < **yabarrik*, [ha:'hu:] ‘that’ < **hahhū*. Most of the phonetic processes had ceased to operate by the time of the Tiberian Masoretes. For instance, pretonic short vowels in open syllables were not lengthened ([jɔ:'ma:ru:] ‘they guarded’ did not shift to [jɔ:ma:ru:]). In such cases, and also in the case of originally long vowels, vowel length was an inherited feature of the language. Some phonetic processes seem to have been still active in the masoretic period. One such process is the general lengthening of all stressed vowels. We know this was a relatively late process (see Khan 1987, 1994a: 133–44).

As a result of the historical background of the Tiberian pronunciation tradition, the vowels *šere* /e/ and *holem* /o/ were always realized as long. The other vowels were realized as either long or short.

In some circumstances there appear to have been differences in duration between stressed vowels that were historically long and those that were historically short. The term “historically long” here refers to vowels that were originally long or that were lengthened by phonetic processes that took place before the masoretic period. “Historically short” refers to vowels that were short or were lengthened by phonetic processes that took place during the masoretic period. In the Tiberian pronunciation tradition, a *šewa* on a letter coming after a historically long vowel was usually silent, e.g. [jo:mri:m] ‘guards’. Such a closed syllable before the main stress could take secondary

stress in the form of an accent: שׁוֹמְרִים [ʃo:m'ri:m]. This implies that the vowel was long enough to accommodate the musical melisma of the accent associated with the secondary stress. Normally, secondary stress was separated from the main stress by an unstressed, buffer syllable, so that the two stress beats did not come together. In a form such as [ʃo:m'ri:m] it appears that the first vowel was lengthened to the extent that it included both the beat of the secondary stress and the unstressed buffer. This would mean that it contained two syllabic peaks: [ʃó:ðmrí:m]. Historically short vowels, on the other hand, could not take the secondary stress in the form of a regular musical accent. When they took secondary stress it was marked by a sign known as a minor *ga'ya*, e.g. [niθħakka'mo:] > [ni:θħakka'mo:] (נִתְחַכְמָה) 'let us deal wisely'. Such cases of *ga'ya* rarely occur immediately before the syllable bearing the main stress since they were not long enough to accommodate both the beat and buffer in contrast to the first vowel in [ʃó:ðmrí:m]. The Arabic transcriptions, moreover, indicate that a vowel with the so-called minor *ga'ya* (i.e. the type found in closed syllables with a historically short vowel) was shorter than one that could take secondary stress in the form of a regular accent (i.e. syllables with a historically long vowel as in [ʃó:ðmrí:m], [ħo:ʔo:'ðo:m] 'the man').

A vocalic *šewa*, which was a historically short vowel, was sometimes lengthened by secondary stress marked by *ga'ya*, e.g. [bana:ħa'lo:] > [ba:na:ħa'lo:] (בְּנַחֲלָה) 'as an inheritance' (Josh. 13:6). There is evidence that also these vowels were not as long as a historically long vowel in an open syllable with secondary stress, e.g. [ħo:ʔo:'ðo:m] > [ħo:ʔo:'ðo:m] 'the man' (see Khan to appear b).

The analysis of the historically long vowel in a closed syllable with secondary stress as having two peaks has implications for the phonemic status of *qameṣ*. It was remarked above that pairs such as [ʔo:ħ'lo:] (אָכַלָה) 'food' vs. [ʔo:ħ'lo:] (אָכַלָה) 'she ate' seem to require us to identify short and long *qameṣ* as two separate phonemes. If the syllable structure of the second word was in fact [ʔo:ħ'lo:], then this would not be a minimal pair proving the phonemic status of the length of *qameṣ*.

There was ambiguity in the syllabic status of some short vowels in open syllables, notably [ɔ] (represented by the *ḥatep qameṣ* sign) in words such as דְּמִי [dɔ'mi:] 'silence', צָרִי [sɔ'ri:] 'balsam', צִפְרִים [ʃippɔ'ri:m] 'birds', כְּתָנֹת [kut-tɔ'no:θ] 'tunics' (Ex. 28:40), הַגְרָנוֹת [haggɔ'no:θ] 'the threshing floors' (Joel 2:24). The vowel [ɔ] in these words was the reflex of an originally short [o] or [u]. The syllable with the short [ɔ] vowel sometimes took secondary stress

and the *ḥateḇ* sign was replaced by an ordinary *qameṣ* in the model Tiberian manuscripts, e.g. [qəðə:'fi:m] (קִדְּשִׁים) > [qə:ðə:'fi:m] (קִדְּשִׁים) 'holy things'. This differs from the occurrence of secondary stress marked by *ga'ya* on a vocalic *šewa* sign, which was not replaced by a full vowel sign, e.g. [ba-na:ḥa'wə:] > [ba:na:ḥa'wə:] (בְּנַחֲלָה) 'as an inheritance' (Josh. 13:6).

Moreover, the writing of ordinary *qameṣ* in place of *ḥateḇ qameṣ* is found in some model Tiberian manuscripts also in a pretonic syllable. The medieval grammarian Ibn Janāḥ refers to the vocalic *šewa* being "lighter" than *ḥateḇ qameṣ* in such words. This implies that there was a difference in length. According to Saadya Gaon (1891: 79), the rules for the occurrence of the apical-alveolar allophone of the Tiberian *reš* treated the word [šə'ri:] 'balsam' as having two syllables. As we have seen, these rules treat a consonant with vocalic *šewa* as belonging to the following syllable.

There is reason to believe, therefore, that in words such as [šə'ri:], [də'mi:] the *ḥateḇ qameṣ* vowel was longer than a vocalic *šewa*. This applies both to cases where the syllable was unstressed and those in which it had secondary stress. This difference in length was sufficient to give the consonant + vowel sequence the status of a independent syllable as reflected by the rules for the distribution of the allophones of Tiberian /r/. We may describe these vowels as half long (CV'), lying in between short vowels (CV) and long vowels (CV:). It appears that a half long vowel could act as a coda of a principal syllable, whereas a short vowel could not.

If the *ḥateḇ qameṣ* was a principal syllable nucleus, then the long and short *qameṣ* in minimal pairs such as [də'-mi:] 'silence' and [də:-mi:] 'my blood' would have to be identified as separate phonemes, since vowel length is the only feature that contrasts them. Since the phonemic contrast is between only two degrees of length, the phonemes could be represented as short /ɜ/ vs. long /ɜ:/.

This could apply in general to cases of *ḥateḇ* vowels that have not been leveled to the normal quality of *šewa* but have a quality close to that of the original short vowel from which they developed. If this is correct, the validity of the aforementioned pairs as proof of phonemic contrasts of short vowels in open syllables would be in doubt, viz. [ʔanij'jə:] 'mourning' vs. [ʔənij'jə:] 'ship', [ḥa'li:] 'ornament' vs. [ḥə'li:] 'illness', [ša'li:] 'go up!' (imperative fem.sg.) vs. [ʃe'li:] 'pestle'. This is because the two members of each pair would have had a different syllable structure. The syllables with [ɔ] and [ɛ] had a quality close to that of the original vowel:⁸ [ʔə'-nij'-jə:] < *ʔniyyā

8. The original quality is preserved in the Babylonian tradition of Hebrew; cf. Yeivin 1985, 2: 876–79.

vs. [ʔanij-ʔo:] ‘mourning’, [ḥoʔ-ʔi:] ‘illness’ < **ḥuly* vs. [ḥaʔli:] ‘ornament’, [ʕe+-ʔi:] ‘pestle’ < **ʕily* vs. [ʕaʔli:] ‘go up!’ (fem.sg.).

6.3. Summary of the phoneme inventory with the known allophones

6.3.1. Consonants

6.3.1.1. Labials

/b/ [b], [v]

/m/ [m]

/p/ [p^h], [f], [p]

/w/ [v], [w]

6.3.1.2. Dentals/alveolars

/t/ [t^h], [θ]

/d/ [d], [ð], [ð]

/t̪/ [t̪]

/s/ [s]

/z/ [z], [z̪](?)

/ʃ/ [ʃ], [z̪]

/ʒ/ /ʒ/

/n/ [n]

/l/ [l]

6.3.1.3. Palatal

/j/ [j], [ʃ]

6.3.1.4. Velars and uvulars

/k/ [k^h], [χ]

/g/ [g], [ʁ]

/q/ [q]

/R/ [R], [r]

6.3.1.5. Laryngeals and Pharyngeals

/h/ [h]

/ʔ/ [ʔ]

/ħ/ [ħ]

/ʕ/ [ʕ]

6.3.2. Vowels

(In the following phonemic notation, /V/ is a phoneme unspecified for length, /Ṽ/ and /V̄/ are phonemes which contain length as a component feature.)

/a/ [a], [a:], [a], [a:]

/ε/ [ε], [ε:]

/ɔ/(?) [ɔ], [ɔ:]

/ṡ/(?) [ṡ:]

/e/ [e:]

/o/ [o:]

/u/ [u], [u:]

/i/ [i], [u:]

/∅/ [∅], [a], [ε], [ɔ], [e], [o], [i], [u]

References

- Abraham ibn Ezra. 1791. *Sefer ha-Moznayim*, ed. Wolf Heidenheim. Offenbach.
- Baer, Seligman, and Hermann L. Strack, eds. 1879. *Die Dikduke Ha-T'amim des Ahron ben Moscheh ben Ascher*, Leipzig.
- Cantineau, Jean. 1950. "Essai d'une phonologie de l'hébreu biblique." *Bulletin de la Société de Linguistique de Paris* 46: 82–122.
- David ben Abraham al-Fāṣī. 1936–45. *The Hebrew–Arabic Dictionary of the Bible Known as Kitāb Jāmi' al-Alfāẓ (Agrōn)*. 2 vols. (Yale Oriental Series, Researches 20–21), ed. Solomon L. Skoss. New Haven: Yale University Press.
- David Qimḥi. 1952. *David Qimḥi's Hebrew Grammar (Mikhlol) Systematically Presented and Critically Annotated by William Chomsky*. New York: Bloch.
- Dukes, Leopold. 1845–46. *Qunṭras ha-masoret*. Tübingen.
- Eldar, Ilan. 1978. *The Hebrew Language Tradition in Medieval Ashkenaz (ca. 950–1350 C.E.)*, vol. 1: *Phonology and Vocalization* (in Hebrew). Jerusalem: Hebrew University.
- . 1980–81. "Hidāyat al-Qāri (the longer Arabic version): A Specimen Text, Critically Edited, with Hebrew Translation, Commentary and Introduction" (in Hebrew). *Lěšonénu* 45: 233–59.

- . 1984. "The Law of *אוי"ה* and *בגדכפ"ת*" (in Hebrew). *Hebrew Union College Annual* 55: א-יד.
- . 1984–85. "The Two Pronunciations of Tiberian *rêš*" (in Hebrew). *Lěšonénu* 48–49: 22–34.
- Garbell, Irene. 1958–59. "The Phonemic State of the šwa, the Ḥāṭēfim and Spirantal *ב ג ד כ פ ת* in Masoretic Hebrew" (in Hebrew). *Lěšonénu* 23: 152–55.
- Grossberg, Menasseh. 1902. *Sefer Yezirah Ascribed to the Patriarch Abraham, with Commentary by Dunash ben Tamim*. London.
- Harris, Zellig. 1941. "The Linguistic Structure of Hebrew." *Journal of the American Oriental Society* 61: 143–67.
- Ibn Janāḥ. 1880. "Kitāb at-takrīb wat-tashīl." In *Opusculs et traités d'Abou 'l-Walīd Merwan Ibn Djanah*, ed. Joseph and Hartwig Dérenbourg, pp. 268–342. Paris: Imprimerie Nationale.
- Khan, Geoffrey. 1987. "Vowel Length and Syllable Structure in the Tiberian Tradition of Biblical Hebrew." *Journal of Semitic Studies* 32: 23–82.
- . 1989. "The Pronunciation of *מדה* before *dageš* in the Medieval Tiberian Hebrew Reading Tradition." *Journal of Semitic Studies* 34: 433–41.
- . 1992. "The Function of the Shewa Sign in Judaeo-Arabic Texts." In *Genizah Research after Ninety Years: The Case of Judaeo-Arabic*, ed. Joshua Blau and S. C. Reif, pp. 105–11. Cambridge.
- . 1994a. "The Historical Background of the Vowel *šere* in some Hebrew Verbal and Nominal Forms." *Bulletin of the School of Oriental and African Studies* 57: 133–44.
- . 1994b. "The Pronunciation of the Verbs *חיה* and *היה* in the Tiberian Tradition of Biblical Hebrew." In *Semitic and Cushitic Studies*, ed. Gideon Goldenberg and Shlomo Raz, pp. 133–44. Wiesbaden: Harrassowitz.
- . to appear a. "The Pronunciation of *reš* in the Tiberian Tradition of Biblical Hebrew." *Hebrew Union College Annual*.
- . to appear b. "The Pronunciation of *šewa* with *ga'ya* in the Tiberian Tradition of Biblical Hebrew." *Vetus Testamentum*.
- Kutscher, Eduard Yehezkiel. 1965. "Contemporary Studies in North-western Semitic." *Journal of Semitic Studies* 10: 21–51.
- Maḥberet ha-Tījān*. 1870. *Manuel du lecteur*, ed. Joseph Dérenbourg. Paris.
- Mann, J. 1931–35. *Texts and Studies in Jewish History and Literature*. 2 vols. Cincinnati: Hebrew Union College Press.
- Mishael ben Uzziel. 1965. *Kitāb al-Kīlāf* (in Hebrew), ed. Lazar Lipschütz. Jerusalem: Hebrew University.

- Morag, Shlomo. 1959–60. “‘The Seven Double Letters כפר”ת בג”ד” (in Hebrew). In *Studies in Honour of N. H. Tur-Sinai* (Publications of the Israel Society for Biblical Research 8), pp. 207–42. Jerusalem.
- Pulgram, Ernst. 1970. *Syllable, Word, Nexus, Cursus* (Janua Linguarum series minor 81). The Hague: Mouton.
- Roman, André. 1983. *Étude de la phonologie et de la morphologie de la koinè arabe*. 2 vols. Aix-en-Provence: Université de Provence.
- Saadya Gaon. 1891. *Commentaire sur le Séfer Yesira ou Livre de la création, par le Gaon Saadya de Fayyoun*, ed. Mayer Lambert. Paris: Bouillon.
- Schreiner, M. 1886. “Zur Geschichte der Aussprache des Hebräischen.” *Zeitschrift für die alttestamentliche Wissenschaft* 6: 213–59.
- Steiner, Richard C. 1993. “Emphatic *p* in the Massoretic pronunciation of ‘appadno (Dan 11:45).” In *Hebrew and Arabic Studies in Honour of Joshua Blau*, ed. M. Bar Asher, Z. Ben-Hayyim, M. J. Kister, A. Levin, S. Shaked, and A. Tal, pp. 551–61. Jerusalem: Tel Aviv University and The Hebrew University of Jerusalem.
- Yeivin, Israel. 1980. *Introduction to the Tiberian Masorah*, trans. and ed. E. J. Revell. Missoula, Mont.: Scholars Press.
- . 1985. *The Hebrew Language Tradition as Reflected in the Babylonian Vocalization* (Texts and Studies 12). 2 vols. Jerusalem: Academy of the Hebrew Language.

Jewish Palestinian Aramaic Phonology

Geoffrey Khan

University of Cambridge

Jewish Palestinian Aramaic (JPA) is the dialect of Aramaic that was spoken and written by Jews in Palestine during the Byzantine period (3rd-7th centuries A.D.) and during the beginning of the Arab period (7th century onwards). By the tenth century it appears that Arabic had replaced Aramaic as vernacular of the Jews of Palestine (Goitein 1966: 198, Friedman 1980: 50–51). By the third century A.D., clear dialectal divisions emerge in Aramaic according to geographical area and confessional community. Among scholars of Aramaic this period is known either as Middle Aramaic (Rosenthal 1939: 104–72) or Late Aramaic (Fitzmyer 1966: 19). JPA is classified as a Western Aramaic dialect, together with the contemporary Aramaic dialects of the Christians and Samaritans.

JPA has been preserved in a number of written sources. These include (i) the Palestinian Talmud, (ii) Palestinian Midrashim (exegetical discussions of the Bible), (iii) texts discussing Jewish Law (*halakha*), (iv) inscriptions, (v) Targums (interpretive translations of the Bible) that were written in Palestine during the Byzantine period, (vi) poetry, (vii) letters and documents preserved on papyrus in Egypt datable to the 5th century, (viii) marriage contracts (*ketubbot*) following the Palestinian tradition of Jewish law written in the Arab period, (ix) magical texts, and (x) marginal notes written in medieval Tiberian Bible codices by the Masoretes. For the bibliography relating to these sources see Sokoloff (1990: 19–28).

One can find minor linguistic differences among some of these sources. It has been customary to refer to the dialect of some of the aforementioned texts, especially types (i)–(iv), as Galilean Aramaic. It would be misleading to assume, however, that there was a geographical dialectal cleavage between the Jewish Aramaic of southern Palestine and that of the Galilee further north. Although the centers of learning and literary activity during the Amoraic period (A.D. 200–500) were in Galilee, where the Rabbinic texts were produced, the inscriptions in JPA datable to this period, the language

of which is very close to that of the Rabbinic texts, have been found throughout Palestine (Sokoloff 1978: 161). In the present state of research it now seems that much of the linguistic variation can be attributed to other factors, such as differences in the chronological period in which the texts were written (Tal 1979) and varying degrees of linguistic conservatism according to the different stylistic registers in which they were written (Greenfield 1974, Yahalom 1993: 332–33).

The language of many of the Rabbinic texts and Targums that were originally written in JPA was corrupted in later manuscripts and printed editions. The late copyists harmonized the language with the Eastern Aramaic of the Babylonian Talmud and the Aramaic of Targum Onkelos (originally written in Palestine but influenced by Babylonian Aramaic in its transmission), with which they were more familiar. For this reason a linguistic description of these must be based on manuscripts which contain the early, uncorrupted form of the text (Kutscher 1976, Sokoloff 1978). Reliable manuscripts of the Rabbinic and Targum texts as well as many of the other aforementioned sources of JPA have been preserved in manuscripts from the Cairo Genizah. This was a medieval repository for discarded manuscripts that was discovered in the Ben-Ezra synagogue of Old Cairo (*al-Fuṣṭāṭ*). Of particular importance for a study of the phonology of JPA are manuscripts with vocalization. Most of these are fragments of a Palestinian Targum that have been discovered in the Cairo Genizah. They are vocalized with either Tiberian or Palestinian vowel signs. Sporadic vocalization is found also in Genizah manuscripts of the Palestinian Talmud and of Aramaic poetry.

7.1. Consonants

The inventory of consonant phonemes (based on Fassberg 1991: 24–25) is seen in Table 7-1.

Each of these phonemes is represented by a separate letter of the Hebrew alphabet. It is not clear whether the primary articulation of /r/ was as an alveolar or a uvular (see below).

7.1.1. /bgdkpt/

The vocalized texts of the Palestinian Targums indicate that the phonemes /bgdkpt/ had both a plosive and a fricative allophone. In manuscripts with Tiberian vocalization, the plosive allophones are marked by the sign *dageš* and the fricative allophone by the absence of *dageš* and sometimes by the sign *raḥē* (Fassberg 1991: 25).

Table 7-1. Jewish Palestinian Aramaic Consonants

Labials	/b/, /p/, /m/, /w/
Dentals/alveolars	/t/, /d/, /s/, /z/, /ʃ/, /n/, /t̪/, /s̪/, /l/, /r/ (?)
Palatal	/y/
Velars	/k/, /g/
Uvular	/q/, /r/ (?)
Pharyngeals	/ʕ/, /ħ/
Laryngeals	/ʔ/, /h/

There is no direct evidence for the precise phonetic realization of the fricative allophones of /bgdkpt/, though it is likely that they were the same as, or close to, their counterparts in the Tiberian pronunciation tradition of Biblical Hebrew and Aramaic, viz. [v], [ɤ], [ð], [χ], [f], [θ] respectively (see chapter 6). The occurrence of the fricative allophones was originally conditioned by the existence of a preceding vowel. As in Tiberian Hebrew, however, due to further phonetic processes that took place before the period in which the vocalized manuscripts were produced (10th–11th centuries A.D.), the fricative allophones are sometimes found also after a consonant, e.g. *ʔayti* ‘he brought’ (Fassberg 1991: 25).

Many JPA texts exhibit the interchange in their orthography of the letters *beth*, which normally represents /b/, and *waw*, which normally represents /w/. When this occurs, either *beth* represents consonantal /w/ or *waw* represents the fricative allophone of /b/. The first of these is far more common than the second. When representing a consonant, the letter *waw* is often written twice to distinguish it from its use as a *mater lectionis* for the vowel /o/. Examples: *lgb* ‘inside’ (= *lgww*, Kutscher 1976: 16), *wrbh* ‘and he became intoxicated’ (= *wrwwh*, *ibid.*), *lmhby* ‘to be’ (= *lmhw*, Friedman 1980: 56), *sytbh* ‘winter’ (= *sytwh*, Naveh and Shaked 1993: 43), *wdlwyy* ‘and of plane-tree’ (= *wdlbyy*, Fassberg 1991: 25). This demonstrates that the fricative allophone of /b/ had the same phonetic realization as the consonant /w/. This may have been [v] or [w]. It was most likely [v], since this was the realization in the Biblical reading traditions in Palestine (see chapter 6).

7.1.2. /s/

The phoneme /s/ is represented either by the letter *sin* or by the letter *samek*, reflecting the merger of the former with the latter, e.g. *סהיד/שִׁהִיד sahed* ‘witness’ (Fassberg 1991: 26; Dalman 1894: 53, 104).

7.1.3. /n/

Final /n/ is often elided after the diphthong /ay/, e.g. *myy* ‘water’ (< *myyn*, Kutscher 1976: 46), *tryy* ‘two’ (< *tryyn*, *ibid.* 26, Svedlund 1974: 31), *bʿyy* ‘asking’ (masc.pl. participle < *bʿyyn*, Kutscher 1976: 44, Svedlund 1974: 31), *hmyy* ‘seeing’ (masc.pl. participle < *hmyyn*, Kutscher 1976: 44).

There is a tendency to append a final *n* to indeclinable words ending in a vowel, reflecting a form of nasalization. Such an addition of final *n* is optional. This is clearly shown by an inscription from Jaffa (Klein 1939: 80, no. 1), where a man is called Ἰουδα in Greek but *ywdn* in Hebrew. In the Greek Septuagint translation of the Bible, the earliest parts of which are datable to the 3rd century B.C., a final *n* is sometimes added to proper names ending in a vowel, e.g. Σαλωμων ‘Solomon’ (< Hebrew *šlmh*). This may be evidence that this phonetic process already existed in the vernacular of the Jews of Palestine and Egypt at that period. Due to the existence of this appended *n* in JPA and its optional omission, an original final *n* in some indeclinable words is dropped. This is attested in a number of proper names, e.g. *ywhnh* < *ywhnn* ‘Johanan’ (Kutscher 1976: 62).

A final *n* is added also to other words. These include, e.g., *sgyn* ‘much’ (< *sgy*, Dalman 1894: 72), *kdown* ‘now’ (< *kdw*, *ibid.* 73), *ʾnty* ‘you (fem.sg.)’ (< *nty*, Friedman 1980: 57), *pwthwn* ‘open’ plural imperative (< *ptwhw*, Yahalom 1993: 334), *pathun* ‘they opened’ (< *ptahu*, *ibid.* 339). Rather than through nasalization of the final vowel, the final *n* in many of these may have a different origin, e.g., by the process of analogy with other words with final *n*. The extent to which the appended *n* occurs, both by final nasalization and by grammatical analogy, depends on the date and literary register of the text. It is less widespread in texts that reflect an earlier phase of JPA than in those reflecting later phases (Tal 1979, Yahalom 1993, Fassberg 1991: 73–74).

In JPA there was a tendency for word-final *m* to shift to *n*. This is attested in proper names, e.g. the place name *Mrwn* (< Biblical *Mrwm* ‘Marom’, Kutscher 1976: 63), and occasionally elsewhere, e.g. *hkyn* ‘clever’ (< *hkym*, Ginsberg 1933: 422). By the phonetic process described in the preceding paragraph the final *n* is sometimes elided, e.g. *Mrw* (< *Mrwn* < *Mrwm*, Kutscher 1976: 61), *Mryh* ‘Maria’ (< *Mryn* < *Mrym* ‘Miriam’, *ibid.*). The shift of final *m* > *n* may have been more extensive in JPA than is revealed by the orthography of the extant texts. It is attested in a number of Hebrew words in texts from Palestine written during the JPA period (Kutscher 1976: 58–60). It is also attested in the modern spoken Arabic dialects of the Levant, e.g. the personal name *Maryen* < *Maryam* (Feghali 1919: 70).

7.1.4. *Pharyngeals* /ħ/, /ʕ/

There is evidence that the pharyngeals /ħ/ and /ʕ/ were weakened in JPA spoken in some regions of Palestine to the extent that they were pronounced as the laryngeals *h* and ‘or even completely reduced to zero. The weakening was apparently brought about by the influence of Greek, particularly in the Hellenized urban centers. In many areas, however, the pronunciation of the pharyngeals was retained. There is no evidence for their weakening in the Palestinian Targum fragments from the Cairo Genizah (Fassberg 1991: 27). Furthermore, scores of place names that were in use in Byzantine Palestine contain to this day the original pharyngeal consonants, e.g. *Ḳirbet Jaʕūn* < *Gʕtwn*, *Yānūḥ* < *Ynwḥ* (Kutscher 1976: 85–86). In modern spoken Western Aramaic, the pharyngeals have been retained.

The shift of the pharyngeal *ḥ* to ‘is attested in several words in the Palestinian Talmud and Midrashim, e.g. ‘*wyḥ* ‘snake’ (< *ḥwyḥ*, Kutscher 1976: 71, Svedlund 1974: 35), ‘*yyb*’ ‘guilty’ (< *ḥyyb*’, Kutscher 1976: 71), and in one instance in the Targum fragments from the Genizah (Fassberg 1991: 65). This shift of the unvoiced pharyngeal *ḥ* to the voiced pharyngeal ‘no doubt reflects a lenition in the articulation, since unvoiced sounds are generally pronounced more fortis than their voiced counterparts. It was, therefore, the first stage in the loss of the pharyngeals.

7.1.5. /r/

We have no direct evidence for the precise phonetic realization of /r/. It should be noted, however, that *a* sometimes shifts to the back vowel *u/o* before /r/, e.g. *šwry* ‘he began’ (< *šari*, Svedlund 1974: 31), *bquryekon* ‘in your cities’ (< *bqaryekon*, Fassberg 1991: 69). This could have resulted from an uvular articulation of /r/ or from a secondary velarization of an alveolar articulation. These are the two phonetic realizations of /r/ in Tiberian Hebrew.

7.2. Vowels

The main source for our knowledge of the vowels of JPA is constituted by the Palestinian Targum fragments from the Cairo Genizah that are vocalized with Tiberian or Palestinian vowel signs (Fassberg 1991: 28–57). The other Palestinian Targums, viz. Targum Pseudo-Jonathan and the Fragment Targum, have Tiberian vocalization in their printed editions, but this is heavily influenced by the vocalization of Biblical Aramaic and Targum Onqelos and does not represent the original vowel system of JPA. Even the vocalization

of the Targum fragments from the Genizah imitates that of the Tiberian tradition of Biblical Aramaic, and so caution is necessary when interpreting the data.

Although the vocalization in the Targum fragments from the Genizah was influenced by the standard Tiberian tradition of Biblical Aramaic, the vowel system of the language of the Targum fragments differed from that of the Tiberian pronunciation tradition. The vocalization of the Targum fragments appears to reflect the existence of five vowel phonemes: /a/, /e/, /i/, /o/, /u/ (Fassberg 1991: 30) and *šewa* vowels, which should perhaps be assigned to the phoneme /Ø/. There are no phonemic contrasts of vowel length.

7.2.1. /a/

In manuscripts with Tiberian vocalization, /a/ is represented by either the vowel sign *qameš* or the sign *pataḥ*. Likewise, in those with Palestinian vocalization it is represented by two vowel signs, corresponding to Tiberian *qameš* and *pataḥ*. Tiberian *pataḥ* and its Palestinian equivalent predominates. This suggests that the phonetic realization of /a/ was closest to that of Tiberian *pataḥ*, i.e. an open front [a]. The confusion of *pataḥ* and *qameš* is attested also in the sporadic vocalizations of the Aramaic of Targum Neophyti and of the Palestinian Talmud (Fassberg 1991: 52).

7.2.2. /e/

In Targum manuscripts from the Genizah with Tiberian vocalization, /e/ is represented by one of the three signs *segol*, *šere*, or *hireq* (Fassberg 1991: 34–38). The most frequently used sign in most contexts is *segol*, suggesting that the phonetic realization was close to that of *segol* in the Tiberian pronunciation tradition, i.e. a half-open front [ɛ]. In non-final, open, stressed syllables, however, only *šere* is used, which implies that in this environment the realization was a higher vowel [e] corresponding to the quality of Tiberian *šere*.

In the Targum fragments from the Genizah, short [i] has shifted to a lower quality and has merged with the phoneme /e/, e.g. *ʿennun* ‘they’ (< *innun*, Fassberg 1991: 35), *lebbeh* ‘his heart’ (< *libbeh*, *ibid.*). The occasional vocalization with Tiberian *hireq* in unstressed closed syllables with an original [i] may reflect a fluctuation in the phonetic realization between [i] and [ɛ]. Alternatively it could be due to the influence of the standard Tiberian tradition of Biblical Aramaic, where short [i] is preserved. The [i] > [ɛ] shift is attested also in the sporadic vocalizations of the Aramaic of Targum Neophyti and

of the Palestinian Talmud (Fassberg 1991: 52, Kutscher 1968: 233). Greek and Latin transcriptions of proper nouns and Hebrew words from the Septuagint (3rd century B.C.) until Jerome (4th–5th century A.D.) reflect the shift [i] > [ε] in closed unstressed syllables. This demonstrates that the shift had taken place in the pronunciation of Aramaic and the sub-standard pronunciation of Hebrew at this period among the Jews of Palestine. In the standard reading tradition of the Bible, these shifts were resisted to a greater extent (Kutscher 1968: 219–27).

In some cases /e/ in an unstressed closed syllable is the reflex of an original short *a* vowel. These can be divided into (i) those that are paralleled by an *a* > *i/e* shift in the equivalent word in a wide range of other Aramaic dialects and represent a shift that took place early in the history of Aramaic, e.g. *dekrin* ‘males’ (< *dakrin*, Fassberg 1991: 35), and (ii) those that are particular to JPA (and occasionally to some other Aramaic dialects of Syria-Palestine) and represent a later extension of this phonetic process to a wider range of words, e.g. *yemma* ‘the sea’ (< *yamma*, *ibid.* 66), *kepna* ‘the famine’ (< *kapna*, *ibid.*).

In late texts written in the Arab period, long *ā* is occasionally represented by the *mater lectionis* *y*, which reflects the raising of the vowel to the region of [ē], e.g. *dmšmšyn* ‘who serve (fem.pl.)’ and *šny’yth* ‘modest (fem.pl.)’ (Friedman 1980: 58) in place of the expected fem. pl. endings *-ān*, *-ātā*. This shift took place most likely under the influence of Arabic, where long [ā] was raised towards [ē] by a process known as *‘imāla*. It is attested also in Christian Palestinian Aramaic texts from the same period (Müller-Kessler 1991: 60).

/e/ in the Targum fragments from the Genizah is sometimes the reflex of an original diphthong *ay*, e.g. *millekon* ‘your words’ (< *millaykon*, Fassberg 1991: 36).

7.2.3. /i/

/i/ is the reflex of [ī] vowels that were long at the time of the [i] > [ε] shift, e.g. *mitu* ‘they died’, *qfila* ‘the dead one’ (Fassberg 1991: 38).

7.2.4. /o/

The vocalized Targum fragments appear to reflect a general shift of short [u] to [o], e.g. *qorbanak* ‘your sacrifice’ (< *qurbanak*, Fassberg 1991: 40), *yeqtol* ‘he will kill’ (< *yiqtul*, *ibid.* 41). The vowel /o/ (< [u]) is occasionally represented with Tiberian *šureq*, *qameš* and even *pataḥ*. Most of these cases, however, can be explained as arising through the influence of the standard

Tiberian vocalization of Biblical Aramaic (ibid. 40–42). Greek and Latin transcriptions of proper nouns and Hebrew words from the Septuagint (3rd century B.C.) until Jerome (4th–5th century A.D.) reflect the shift [u] > [o] in closed unstressed syllables (Kutscher 1968: 219–27).

/o/ before a labial consonant is sometimes the reflex of an original short *a*. This is represented by the *mater lectionis* *w*, e.g. *šwbh* ‘Sabbath’ (< *šabbā*, Kutscher 1976: 17), *ʿwp* ‘also’ (< *ʿap̄*, ibid.), *šwbʿyn* ‘seventy’ (< *šabʿīn*, Friedman 1980: 57), *rwbyn* ‘large (masc.pl.)’ (< *rabbīn*, ibid.). In the vocalized fragments of the Targum from the Genizah this vowel is [o] or [u], which are both allophones of /o/ (see above), e.g. *gobrin* ‘men’ (< *gabrin*, Fassberg 1991: 69), *šubʿa* ‘seven’ (< *šabʿa*, Fassberg 1991: 69). This shift occurs occasionally before *r* and *l*, e.g. *pwlg* ‘half’ (< *palga* Friedman 1980: 58), *šwry* ‘he began’ (< *šari*, Svedlund 1974: 31), *bquryekon* ‘in your cities’ (< *bqaryekon*, Fassberg 1991: 69).

In some cases /o/ is the reflex of an original diphthong [aw], e.g. *tora* ‘the ox’ (< *tawra*, Fassberg 1991: 41), *ʿoledt* ‘you begat’ (< *ʿawledt*, ibid.).

7.2.5. /u/

/u/ is the reflex of [ū] vowels that were long at the time of the [u] > [o] shift, e.g. *ydunun* ‘they will judge’, *sebu* ‘old age’ (Fassberg 1991: 42).

7.2.6. /Ø/ ? (*šewa*)

Short vowels in open unstressed syllables after a non-guttural consonant are generally represented by the *šewa* sign in Targum fragments with Tiberian vocalization. We have little external evidence for vowel length differences in JPA. Let us assume, however, that vowels in open syllables that are regularly represented by vowel signs rather than the *šewa* sign (e.g. *napeq* ‘going forth’) were long, as was the case in Tiberian Hebrew. If this were the case, then vowels marked by the *šewa* sign would be distinguished from other vowels in open syllables by their shortness. It seems, moreover, that such short vowels in open syllables would not contrast phonemically with zero, so perhaps they should be analyzed as allophones of the phoneme /Ø/, as is the case with *šewa* vowels in Tiberian Hebrew phonology. Occasionally the *pataḥ* [a] sign is used where one would expect *šewa*, suggesting that vocalic *šewa* was pronounced with the quality of [a], which was its normal phonetic realization in the Tiberian pronunciation tradition (Fassberg 1991: 47) and is attested also in some Hebrew manuscripts with Palestinian vocalization.

The interchange of the Tiberian signs *šewa* and *pataḥ* is found also in the vocalization of poetry written in JPA (Yahalom and Sokoloff 1995 Introduction). Furthermore, when two vocalic *šewas* follow one another in the Targum manuscripts from the Genizah, the second *šewa* is elided and the first is represented by an *a* vowel sign, e.g. *labsar* ‘to the flesh of’ (Fassberg 1991: 107). Conversely, the *šewa* sign is sporadically used to represent an [a] in a closed syllable (ibid. 33). Before /y/ vocalic *šewa* is often represented by *hireq* [i], which also corresponds to the Tiberian pronunciation tradition. In manuscripts with Palestinian vocalization an *e* vowel sign is often used where vocalic *šewa* would occur in Tiberian vocalization (ibid. 46). The practice of representing vocalic *šewa* with an *e* vowel sign is found also in Hebrew manuscripts with Palestinian vocalization. It is possible, therefore, that the Tiberian and Palestinian reading traditions of Hebrew influenced the scribes who vocalized the Aramaic texts, and it is difficult to establish with any certainty the authentic quality of the vocalic *šewa* in the Aramaic.

Vocalic *šewa* is, occasionally, represented by a *mater lectionis*. In Targum texts with Palestinian vocalization, *mater lectionis yod* is found in this function (Fassberg 1991: 63). In some amulets and inscriptions in JPA, vocalic *šewa* before a labial is represented by *mater lectionis waw*. This is attested in the word *šwmyh* ‘heaven’ (< *šamayya*, Naveh 1978: 40–42, 106–9; Naveh and Shaked 1985: 50–54, 82–85). The [a] > [o] shift before labials, therefore, operated on short vowels in closed and open syllables alike. The existence of the vowel [o] shows that the quality of the vowel before the shift must have been [a], i.e., that of the Tiberian *šewa*. It is not clear whether this can be taken as evidence that short vowels in open unstressed syllables in general were leveled to the quality of [a], as in the Tiberian tradition.

7.3. Diphthongs

In most environments, the diphthong /ay/ does not contract. In unvocalized texts the existence of the diphthong is shown by the many cases where the [y] is spelled with double *yy*, which indicated that the letter was to be pronounced as consonantal /y/ and not as a vowel: *byyty* ‘my house’, *‘yynh* ‘the eye’, *ḥyylyh* ‘the strength’ (Kutscher 1976: 19). /ay/ contracts to /e/ before pronominal suffixes on masc. plural and dual nouns, e.g. *bnek* ‘your children’ (Fassberg 1991: 58).

The reflex of an original diphthong *aw*, on the other hand, is generally /o/ in JPA, e.g. *tor* ‘ox’ (< *tawr*, Fassberg 1991: 58).

References

- Dalman, Gustaf. 1894. *Grammatik des Jüdisch-Palästinischen Aramäisch*. Leipzig: Hinrichs.
- Fassberg, Steven E. 1991. *A Grammar of the Palestinian Targum Fragments from the Cairo Genizah* (Harvard Semitic Series 38). Atlanta: Scholars Press.
- Feghali, Michel T. 1919. *Le parler de Kfar 'Abîda (Liban-Syrie)*. Paris: Geuthner.
- Fitzmyer, Joseph A. 1966. *The Genesis Apocryphon of Qumran Cave I: A Commentary* (Biblica et Orientalia 18). Rome: Biblical Institute Press.
- Friedman, Mordechai A. 1980. *Jewish Marriage in Palestine*. Tel-Aviv: Tel-Aviv University.
- Ginsberg, H. Louis. 1933. "Zu den Dialekten des Talmudischen-Hebräischen." *Monatsschrift für Geschichte und Wissenschaft des Judentums* 77: 413–29.
- Goitein, Shlomo D. 1966. "Four Old Marriage Contracts from the Cairo Genizah" (in Hebrew). *Lěšonénu* 30: 197–216.
- Greenfield, Jonas C. 1974. "Standard Literary Aramaic." In *Actes du premier congrès international de linguistique sémitique et chamito-sémitique* (Janua Linguarum Series Practica 159), ed. André Caquot and David Cohen, pp. 280–89. The Hague: Mouton.
- Klein, Samuel. 1939. *Sefer ha-Yiššuv*, vol. 1. Jerusalem: Dvir.
- Kutscher, Eduard Y. 1976. *Studies in Galilean Aramaic*, trans. Michael Sokoloff. Ramat-Gan: Bar-Ilan University.
- . 1968. "Articulation of the Vowels *u*, *i* in Transcriptions of Biblical Hebrew, in Galilean Aramaic, and in Mishnaic Hebrew" (in Hebrew). In *Benjamin de Vries Memorial Volume*, ed. E. Melamad, pp. 218–51. Jerusalem: Tel-Aviv University and Stichting Fronika Sander Fonds.
- Müller-Kessler, Christa. 1991. *Grammatik des Christlich-Palästinisch-Aramäischen*, part 1. Hildesheim: Olms.
- Naveh, Joseph. 1978. *On Stone and Mosaic: The Aramaic and Hebrew Inscriptions from Ancient Synagogues* (in Hebrew). Tel-Aviv: Israel Exploration Society.
- Naveh, Joseph, and Shaul Shaked. 1985. *Amulets and Magic Bowls: Aramaic Incantations of Late Antiquity*. Jerusalem: Magnes.
- . 1993. *Magic Spells and Formulae: Aramaic Incantations of Late Antiquity*. Jerusalem: Magnes.

- Rosenthal, Franz. 1939. *Die aramaistische Forschung seit Th. Nöldekes Veröffentlichungen*. Leiden: Brill.
- Sokoloff, Michael. 1978. "The Current State of Research on Galilean Aramaic." *Journal of Near Eastern Studies* 37: 161–67.
- . 1990. *A Dictionary of Jewish Palestinian Aramaic of the Byzantine Period*. Ramat-Gan: Bar-Ilan University.
- Svedlund, Gerhard. 1974. *The Aramaic Portions of the Pesiqta de Rab Kahana* (Acta Universitatis Upsaliensis). Uppsala.
- Tal, Abraham. 1979. "Layers in Jewish Palestinian Aramaic" (in Hebrew). *Lěšonénu* 43: 165–84.
- Yahalom, Joseph. 1993. "Verbal Suffixes in Jewish Palestinian Aramaic" (in Hebrew). In *Hebrew and Jewish Studies in Honour of Joshua Blau*, ed. M. Bar Asher, Z. Ben-Hayyim, M. J. Kister, A. Levin, S. Shaked, and A. Tal, pp. 331–40. Jerusalem: Tel-Aviv University and the Hebrew University of Jerusalem.
- Yahalom, Joseph, and Michael Sokoloff. 1995. *Aramaic Poetry of the Byzantine Period from Eretz Israel* (in Hebrew). Jerusalem: Israel Academy of Sciences and Humanities.

This page intentionally left blank.

Old Aramaic Phonology

Stanislav Segert

University of California, Los Angeles

8.1. Terms and problems

Under the heading “Old Aramaic” the basically uniform, but not monolithic language preserved in sources composed in the first millennium B.C. is dealt with here. “Biblical Aramaic,” the language of texts mostly from the 6th, 5th and 2nd centuries B.C., is one part of Old Aramaic.

The terminology for stages and dialects of Old Aramaic is not uniform. Here the term “Old Aramaic” is used for the entire first millennium B.C. In some other publications this term is applied only to the older stage, until 612 B.C. approximately; for this stage the term “Early Aramaic” is used here.

The use of Aramaic as the official language in a great part of the Persian Empire is expressed in the terms “Imperial Aramaic,” calqued from German *Reichsaramäisch*, or more generally, “Official Aramaic.” From a linguistic viewpoint “Biblical Aramaic” is a part of this stage. As the Aramaic texts from the second half of the first millennium B.C. reflect mostly the uniform language of the Persian Empire even after its termination in the 4th century B.C., the period of “Imperial Aramaic” can be extended to the 2nd century B.C., before the splitting of the Aramaic languages into Western and Eastern groups.

The notion of “phonology” in its narrower sense is applicable for the written documents, the phonemes of which have to be reconstructed. But the pronunciation of the Aramaic passages of the Bible has been preserved also by oral tradition, which could be influenced by the later development of Aramaic and its related languages. Thus some phonetic observations can be made in this survey.

The overwhelming majority of Old Aramaic texts are written in an alphabet taken over from Phoenician. A few texts and some words are attested in

ABBREVIATIONS: BA, Biblical Aramaic; DP, Demotic Papyrus; EA, Early Aramaic; IA, Imperial (Official) Aramaic; OA, Old Aramaic; SA, Samalian Aramaic.

the Mesopotamian cuneiform syllabary. An important manuscript on papyrus, Amherst Egyptian 63 (containing mostly poetry), is written in Egyptian demotic script. This important text is being edited by Richard C. Steiner; he kindly provided relevant data which are incorporated in the text of this chapter.

Aramaic was used for communication between peoples of different language backgrounds, and also in areas where other languages were more common. In the Assyrian and Babylonian Empires Aramaic connected the ruling class speaking Akkadian with the common Aramaic people. For contact with Hebrew speakers cf. 2 Kings 18:26. Judeans returning from the Babylonian exile brought with them the Aramaic language. In the multilingual Persian Empire Aramaic was introduced as official language in large areas of Western Asia and also in Egypt. After the demise of the Persian Empire in the 4th century B.C., Aramaic was still used in many countries.

The combination of data from Aramaic texts along with the traditional pronunciation, and some features of other Aramaic and related languages, make it possible to reconstruct Old Aramaic phonology, even if this is not always perfectly reliable.

8.2. Sources

When Emil Kautzsch published his German grammar of BA in 1884, very few other Old Aramaic texts were known. Since that time hundreds of inscriptions and papyri have been found, mostly by excavations in various Near Eastern countries.

Aramaic biblical texts in the books of Ezra and Daniel represent about 1% of the Old Testament canon. Comparison of complete Masoretic texts written around 1000 A.D. with fragments from the Qumran area near Jericho shows that for 1,000 years these texts were transmitted with few changes in the consonantal writing. Vowel signs were added in the last centuries of the first millennium A.D.; they are completely accessible according to the Tiberian system, with some sections also in the older Babylonian vocalization.

The archaic dialect of the kingdom of Sam'al ('North') is represented by three inscriptions from the 8th century B.C. found in the ruins of the capital of Ya'udi (near Zincirli in southern Turkey).

Early or Ancient Aramaic texts on stone, from the 9th–7th century B.C., were preserved in Syria. Texts from the later part of this period were written

on clay tablets and ostraca (sherds) in Assyria (northern Iraq).

Texts from the Imperial Aramaic stage, about the 7th–2nd century B.C., were found in many countries, from western Asia Minor to the Indus Valley, from the Caucasus to Upper Egypt. They were written on stone and sherds. A great number of papyri were preserved in Egypt, mostly from the Jewish military colony on the Nile island Jeb (Elephantine) near Seven (Aswan).

The clay tablet in syllabic cuneiform script found in Uruk in Lower Mesopotamia shows, like some other texts from Mesopotamia, some features of the Eastern Imperial Aramaic dialect.

A large collection of Aramaic texts written in Egyptian demotic script is preserved on Papyrus Amherst Egyptian 63 in the Pierpont Morgan Library in New York. After short samples were published by Raymond Bowman in 1944, Charles F. Nims and Richard C. Steiner have published since 1983 several cultic poetic texts. Steiner is now preparing a complete edition of the papyrus. This papyrus scroll found at Thebes in Egypt was written probably in the early Ptolemaic period, in the third century B.C., but many poems, like that parallel to biblical Psalm 20 or that on an Assyrian king, point to older traditions. This Upper Egyptian text exhibits features of Western Aramaic, including some which were previously thought to be Late Aramaic innovations.

Quotation from sources in the alphabetic script are presented here in italics, those from cuneiform syllabaries also in italics, with dashes between transcriptions of syllabic signs. The quotations from the text in the demotic script are marked DP (Demotic Papyrus).

8.3. Kinds of phonological data

Ancient phonological features are represented by phonemic signs and graphic devices in ancient and medieval texts. This evidence is supplemented by phonetic observations on the traditional pronunciation patterns of Biblical Aramaic.

The alphabetic script of Phoenician origin has 22 graphemes. In principle each of them indicated one consonant phoneme.

As in Early Aramaic the number of consonant phonemes was more than 22, some letters had to indicate more than one consonant. E.g., the letter *z* was used both for the sibilant /z/ and interdental /d̪/; cf. § 8.4.1.

Since the 9th century B.C. some consonant letters were used also for indication of vocalic phonemes. The letter *w* could indicate the bilabial

consonant /w/, the long vowel /ū/, or the diphthong /au/. Similarly the letter *y* served to indicate the consonant /y/, the long vowel /ī/, or the diphthong /ai/. Also the letters *h* and *ʾ* were used for indicating vowels, mostly long /ā/.

In the late centuries of the first millennium A.D. specific graphic signs for vowels—and accents—were introduced into Biblical Aramaic texts. The older Babylonian system used signs above consonant letters; the Tiberian system put most signs beneath the letters.

The gemination of consonants is marked only in the Biblical Aramaic of the Tiberian tradition.

The cuneiform syllabary indicated by its vocalic components vowels—except /o/, often also their length. However, the rendering of consonant phonemes by consonant components is not always exact.

The demotic script is able to distinguish some phonemes which in the Aramaic alphabet are rendered by a common grapheme. The uvular *ḥ* is distinguished from pharyngeal *ḥ*, e.g. in *ḥ.mr^m* ‘wine’ versus *ḥ.m.^m* ‘venom’ (cf. Ugaritic *ḥmr* ‘wine’, *ḥmt* ‘venom’). The voiced uvular *ḡ* is indicated in DP by *ḥ* or *ḡ*, e.g. *.rḡ.^m* /ʾarḡ-/ ‘earth’.

The uninterrupted tradition of reading Biblical Aramaic texts provided by vocalic and accentual signs in the late first millennium B.C. allows phonetic observation about the pronunciation of sounds indicated by graphemes and additional signs. The older Babylonian tradition was preserved only in Yemen; the younger tradition fixed in Tiberias is commonly employed.

These traditions as well as their graphic bases were influenced by later commonly spoken Aramaic dialects and by the traditional pronunciation of Biblical Hebrew.

8.4. Consonantal phonemes: Inventory and systemic changes

The number of consonant phonemes in EA was larger than in IA. Several allophones developed in BA.

In Table 8-1 this situation is indicated by different graphics: graphemes are in angle brackets, their reconstructed pronunciation within slants, the traditional pronunciation presently observed within square brackets.

8.4.1. Interdentals

The interdentals were in the Early Aramaic texts indicated by letters corresponding to the equivalent words in Canaanite languages, Phoenician and Hebrew: /t/ by *š*, e.g. *šlš* /t-l-t/ ‘three’, cf. Hebrew *šlš*; /d/ by *z*, e.g. *zḥb* /d-h-b/ ‘gold’, cf. Hebrew *zḥb*; /z/ (/t/) by *š*, e.g. *šby* ‘gazelle’, cf. Hebrew *šby*.

Table 8-1. Old Aramaic Consonants

	Early Aramaic			Imperial Aramaic			Biblical Aramaic allophones		
	Emphatic	Voiced	Unvoiced	Emphatic	Voiced	Unvoiced	Emphatic	Voiced	Unvoiced
Laryngeals		<ʾ>	<h>		<ʾ>	<h>			
Pharyngeals		<ʾ>	<ḥ>		<ʾ>	<ḥ>			
Uvulars		/ġ/	/ḥ/				[ġ]		[k]
Velars	<q>	<g>	<k>	<q>	<g>	<k>			
Palatovelars			<š>			<š>			
Sibilants	<š>	<z>	<s>	<š>	<z>	<s>			[ś]
Interdentals	/z (ṭ)/	/d/	/t/				[d]		[t]
Dentals	<ṭ>	<d>	<t>	<ṭ>	<d>	<t>			
Labials			<p>			<p>			
Bilabials							[b]		[p̄]
Linguals		<l r>			<l r>				
Nasals		<n m>			<n m>				
Semivowels		<y w>			<y w>				

In the Tell Fekheriye inscription (9th century B.C.) the interdental /t/ is indicated by the phonetically close sibilant letter *s*, e.g. *y-s-b* /y-ṭ-b/ ‘to sit’, while this word is written in other Early Aramaic texts as *y-š-b*, cf. Hebrew and Phoenician *y-š-b*.

The consonant going back to Proto-Semitic **ḏ*—or perhaps originally lateralized **ḏ^l*—is indicated in EA by the letter *q* which is by its place of articulation close to /ġ/; e.g. *ʾrq* /ʾ-r-ġ/ ‘earth’ (cf. Arab. *ʾ-r-ḏ*, Phoen. and Hebr. *ʾ-r-š*), *qmr* /ġ-m-r/ ‘wool’ (cf. Hebr. *š-m-r*).

In IA this consonant indicated in EA by *q* changed into the pharyngeal ʿ; e.g. *ʾr*, *ʾmr*. For another interpretation of these facts, based on the demotic text, see Steiner (1991a: 1499–1501).

The interdentals changed in IA into dentals; new forms of words listed above were written *tl̄t*, DP *t.rt.^m*; *dḥb*, DP *t.h.b.^m*; *ṭby*.

8.4.2. Other consonants

After vowels some simple plosive consonants, /t/, /b/, /p/, developed in later BA tradition spirant position variants (allophones), [ḡ] (cf. /ḡ/), [ḵ] (cf. Arab. ḥ), [ḏ] (cf. /ḏ/), [ṭ] (cf. /t/), [ḃ] (cf. English [v]), [ḗ] (cf. Arab. f).

The dot above the left side of the letter š in BA texts with Tiberian signs indicates the pronunciation as [ṣ̌] (šin).

8.5. Vocalic phonemes: Inventory and systemic changes

All vowels are indicated in BA texts from about 1000 A.D. These traditions reflect to a large extent the IA pronunciation, but they were affected by later commonly spoken Aramaic dialects and by the traditional pronunciation of Biblical Hebrew. For reconstruction of EA vocalism, indication of vowels by consonant letters and evidence from non-alphabetic sources can be used.

The Tiberian vocalization system is exact, but it applies the patterns of the traditional pronunciation of Hebrew to Aramaic.

The Babylonian vocalization system is simpler and closer to the original BA pronunciation.

The consonant letters y, w, h, and ' were used for long vowels.

In the following survey the vowels according to Babylonian and Tiberian systems are presented:

Babylonian system: *i e a o u* short and long; reduced mixed vowel ^ə

Tiberian system: the same vowels, minus long /ā/, and additional vowels: *æ* and *ā*, short, long, reduced; reduced ^a

Vowel signs are identical for short and long vowels. Long vowels are often indicated by consonant letters which were introduced already in EA and IA stages.

In the Uruk incantation text, which probably reflects the pronunciation of the early IA stage, qualities and quantities of vowels are indicated by vocalic components of cuneiform syllabic signs.

The vowel /o/ is indicated by -u- components; cf. *ti-ḥu-ú-tú* /tiḥōt-/ , cf. BA *t^əḥōt* 'below'.

Length is indicated, but not consistently, by repeating vocalic elements; *qu-ú-mi-* /qūmī-/ 'stand up!' (fem.); but *ša-am-lat* /šamlat/, cf. Arab. *šamlat*-, Hebr. *šimlat* < *šam- 'cloak, clothing'.

The most reliable indicators of vowels in the demotic papyrus are the signs for /w/ and /y/ (including multiconsonantal signs like *sw* and *ty*). In addition to their consonantal value, these signs also have vocalic value, being used to represent high back and high front vowels, respectively.

It can be supposed that the inventory of vowels was smaller in EA and older stages of IA.

Long vowels /ō/ and /ē/ developed mostly as diphthongs /aw/ and /ai/ were monophthongized (cf. § 8.7.2).

Long /ā/ remained basically unchanged also in the Babylonian tradition of BA. In the late Tiberian tradition of BA it is pronounced like Hebrew more closed $\bar{a}$.

8.6. Conditioned consonant changes

8.6.1. By position in word or syllable

The glottal stop ' was elided at the end of words and syllables; e.g. IA *šgy* 'BA *šaggi*('), and IA *šgy*/*šaggī*/ < **-i*'; IA (-)*m'mr*, (-)*mmr*; BA *mē*(')*mar*, *mēmar* /*mēmar*/ < **mi'm*- 'to say'. In the demotic papyrus, elision of ' is also attested as a sandhi phenomenon, e.g., *m̄n.n.nty^m ety^m* [manantī 'attī].

If the postpositive article /-ā/ developed from an original deictic element **ā*, the elision of final glottal stop could be supposed: **-ā* > **-ā'* > /-ā̄/. But the letter ' can be considered a device for indicating the long vowel /-ā/.

The elision of *h* sporadically attested for suffixed pronouns in IA is rather common in the DP, e.g. *mšhṭn^m* 'their sin'. The feminine marker **-t* is elided in the absolute state: **millat*- (cf. construct state *millat*) > *millā(h)* /*millā*/ 'word'; **malkūt*- (cf. constr. *malkūt*) > *malkū* 'kingdom'.

The final consonant of pronouns and perfect forms of 2nd and 3rd person pl. masc. was *-m* in IA and Ezra, but *-n* in the younger BA, in Daniel.

In the DP, final *n* preceded by the diphthong *ay* may be deleted as later in Galilean Aramaic, e.g., *k.py^m* 'overturned' (masc. pl.) and *b.y.^m* 'among'.

8.6.2. By combination of consonants

Assimilation and dissimilation occur mostly at nasals and liquids.

Regressive total assimilation of the nasal /n/ to the immediately following consonant is very frequent, but forms without assimilation are attested. Cf. DP *m̄p.kw* 'take out', EA *ypq* 'he goes out', and IA *tnpq* 'she goes out'; BA *yippel* 'he falls down'; IA *ttn* and *ttn*, BA *tintēn* 'you give' (2nd person sg. masc.). Such assimilation can be observed on the verb *l-q-ḥ* 'to take': EA *yqh* and *ylqh* 'he takes', IA *lqh* 'I take'.

Progressive assimilation is attested for the verb *s-l-q* 'to get up': EA *ysq* 'he gets up', BA *hassíqū* 'they brought up'.

Geminated consonants can be dissolved into *n* + simple consonant: **has-sāqā* > IA *hnsqh*, BA *hansāqā* 'to bring up'; **ha-ālā* (inf. Haf'el of 'l-ḥ; cf.

hæ'ālā BA) > *han'ālā* 'to introduce'; DP *hnḥ.rw hnḡlw* 'they were brought in'. Some forms with *n* + consonant in late IA and in BA may be the result of this dissimilation; cf. *ttn*, *tnnt*, *tinten*.

Dissimilative tendencies were active in replacing consonants close in their articulation place by consonants more distant. Instead of original **b-n*, preserved in plural *bnn* (cf. Hebr. and Phoen. *bn*), Aram. has *br* in EA and IA, *bar* in BA for 'son'; also *brh* 'daughter', pl. *bnt*. Original **n*, preserved in the ordinal numeral IA *tnyn*, BA *tinyānā(h)* (fem.) 'second' (cf. Ugar. *tn*, Hebr. *šny(m)*, Phoen. *šnm* 'two' was replaced by *r* in the cardinal numeral 'two'; IA *tryn*, BA *tartēn* (fem.) DP *ṭry^m*.

Liquids also participate in metathesis. Cf. IA (and BA) *tr* 'gate'; the original sequence is preserved in Ugar. *tḡr* and Hebr. and Phoen. *šr*; cf. SA *lgry* (du. constr.) 'feet', DP *r.kṛyk^m* (*lgryk*) 'your feet' and IA *rgly* 'my foot'.

If two emphatic consonants are together in one root, one of them can be dissimilated: IA *kšph* 'his anger', original *q* preserved in BA *q^ʕšaḗ*; *q-t-l* 'to kill' in SA and RA, *q-t-l* in RA and IA.

8.6.3. By contacts with vowels

The semivowel *y* can be elided between two long vowels; cf. IA *qym* /qāyēm/ and BA *qā'ēm* 'standing up'.

8.7. Conditioned vowel changes

Short vowels in open syllables were reduced or elided in later IA and BA. The transition can be observed in the cuneiform Uruk text: *ga-ba-ri-e* and *ga-ab-ri-e* 'men'. Cf. *š^ʕnaṭ* (constr.) < **š^ʕanat* 'year (of)'; *malk^ʕtā* < **malkatā* 'the queen'; *kāt^ʕbā* 'writing' (fem. partic.) < **kātibat*; *yišk^ʕnūn* 'they dwell'; elision: *ḥaṭmah* < **ḥatam-ah* 'he sealed it (fem.)'.

Primary vowels change into secondary ones in closed syllables: *i* – *e*, e.g. BA *s^ʕgid* 'he did homage' – *š^ʕleṭ* 'he ruled'; *ī* – *ē*, e.g. BA *h^aqīm* and *h^aqēm* 'he set up'; *u* – *o*, cf. BA *gubbā* (determinate) and *gōb* (absolute) '(the) pit'.

In BA the vowel /a/ in a closed unstressed syllable can change into /i/, e.g. **baśrā* (cf. Hebr. *bāśār*) > *bīśrā* 'the flesh'.

New vowels are inserted to avoid clusters of consonants: in the segolate nouns, e.g. **malk-* > *mælæk* 'king', **-lm-* > *š^ʕlem* 'statue'; in the 1st person sg. masc. of perfect: **šam't-* > *šim'eṭ* 'I heard'.

Similar vowels in adjacent syllables can be dissimilated: **'abū-hū* > *'abūhī* 'his father'.

Laryngeals, pharyngeals, and *r* tend to change the adjacent vowels in the direction toward /a/: **sāpīr* > *sāpār* ‘writing (partic.), scribe’; **kāhin-ā* > *kāl^anā* ‘the priest’; **i* – *æ* in the infinitive of the *miqtal* pattern (cf. *mišbaq* ‘to let’) *mæ’baq* ‘to make’. In the Tiberian tradition *a* is inserted between long vowel and the final pharyngeal: *rūaḥ* ‘spirit’, *tērōa* ‘she crushes’.

8.7.1. Compensatory lengthening

Vowels are lengthened as compensation for loss of “weak” consonants.

Compensatory vowel lengthening occurs after the loss of glottal stop: **riʕ* > BA *rē(ʕ)š*, cf. IA *rš* ‘(the) head’; **ba-ʕar* > BA *bā(ʕ)tar* ‘after’, BA *bāl^ʕrāk* ‘after you’.

As compensation for simplification of the doubled consonant the preceding vowel is lengthened: **min-ʕarʕā* > *miʕ-ʕarʕā* > BA *mēʕarʕā* ‘from the earth’; **-hh-* > BA *mitbāhal* ‘frightened’; **-rr-* BA > *bārik* ‘he blessed’, BA *tārā* ‘gate-keeper’.

8.7.2. Monophthongization

Diphthongs /aw/ and /ai/ were monophthongized in IA and BA, into /ō/ and /ē/ respectively: **yawm* > IA *ywm* and *ym*, BA *yōm*; IA *byt*, BA *baytā* ‘the house’; IA *byt*, BA *bē(y)t* (constr.) ‘house (of)’.

8.8. Other phonological features

8.8.1. Word stress

This suprasegmental feature can be observed directly in BA texts provided with accent signs and in their traditional pronunciation. For older stages only inferences supported by reconstructions and analogies can be delivered.

In the Masoretic accentuation of BA (corresponding apparently to Hebrew accentuation) most words have stress on the last syllable. The following types of words have accent on the last but one syllable: segolate nouns and verbal forms with anaptyctic vowel in the last syllable, e.g. *māʕlæḵ* < **mark-* ‘king’, *šāmæt* < **šāmt(u)* ‘I set’; verbal forms with affirmatives ending on long vowel, e.g. *ʕkūlī* ‘eat!’ (fem.), *yē(ʕ)ḥādū* ‘may they perish’, *š^ʕʕlnā* ‘we asked’; weak verbs with long vowel in the 3rd sg. fem. perfect, *sāpāt* ‘she was fulfilled’; words going back to (adverbial) accusatives: *ʕllā* ‘above’; words provided with suffixed pronouns consisting of consonant + vowel, e.g. *ʕbdōhī* ‘his servants’, *haʕlnī* ‘bring me in!’.

Most of the types just given point to original accentuation on the last but one syllable. After short end vowels indicating cases or persons, perhaps partially still preserved in earlier IA—cf. (*mi-in*) *ig-ga-ri* ‘(from) wall’, *ḥa-al-li-tú* /*ʕallítu*/ ‘I entered’, in the Uruk text—the originally last but one syllable became the last one, but the accent remained on it.

8.8.2. Syllables

Quantities of vowels and accents can be exactly observed only in the Tiberian tradition of BA. Some of these features are the results of a relatively late development.

Open syllables with long vowels (C $\bar{V}$) appear as stressed and unstressed. Only rarely open syllables with short vowels (CV) are attested. They changed into syllables with reduced vowels (C^v).

Closed syllables with short vowels (CVC) occur with or without stress, those with long vowels (C $\bar{V}$ C) with stress, thus mostly at the end of words.

These observations may be applied for ostensible consonant clusters at the end of words: cf. ^a*[baḏ | t^ə]* ‘you made’ (sg. masc.)

Selective Bibliography

- Abou-Assaf, Ali, Pierre Bordreuil, and Alan R. Millard. 1982. *La statue de Tell Fekherye et son inscription bilingue assyro-araméenne*. Paris: Recherche sur les civilisations.
- Bauer, Hans, and Pontus Leander. 1927. *Grammatik des Biblisch-Aramäischen*. Halle/Saale: Niemeyer.
- . 1929. *Kurzgefasste biblisch-aramäische Grammatik*. Halle/Saale: Niemeyer.
- Degen, Rainer. 1969. *Altaramäische Grammatik der Inschriften des 10.–8. Jh. v. Chr.* Wiesbaden: Deutsche Morgenländische Gesellschaft.
- Dion, Paul-Eugène. 1974. *La langue de Ya’udi*. Waterloo, Ontario: Corporation for the Publication of Academic Studies in Religion in Canada.
- Drijvers, H. J. W. 1973. “Aramaic.” In *A Basic Bibliography for the Study of the Semitic Languages*, ed. J. A. Hospers, vol. 1, pp. 290–335. Leiden: Brill.
- Fitzmyer, Joseph A., and Kaufman, Stephen A., eds. 1992. *An Aramaic Bibliography*, part 1, *Old, Official, and Biblical Aramaic*. Baltimore: The Johns Hopkins University Press.
- Kautzsch, Emil. 1884. *Grammatik des Biblisch-Aramäischen*. Leipzig: Vogel.

- Kutscher, Eduard Yechezkel. 1970. "Aramaic." In *Current Trends in Linguistics*, ed. Thomas A. Sebeok, vol. 6, *Linguistics in South West Asia and North Africa*, pp. 347–412. The Hague: Mouton.
- Leander, Pontus. 1928. *Laut- und Formenlehre des Ägyptisch-Aramäischen*. Göteborg (repr. Hildesheim: Olms, 1966.).
- Rosenthal, Franz. 1939. *Die aramaistische Forschung*. Leiden: Brill (repr. 1964).
- . 1961. *A Grammar of Biblical Aramaic*. Wiesbaden: Harrassowitz.
- Segert, Stanislav. 1956. "Mluvnice aramejštiny" (A grammar of Aramaic). In Otakar Klíma and Stanislav Segert, *Mluvnice hebrejštiny a aramejštiny* (A grammar of Hebrew and Aramaic), pp. 12–17, 20–21, 237–302. Prague: Nakladatelství Československé akademie věd.
- . 1975. *Altaramäische Grammatik*. Leipzig: Enzyklopädie (repr. 1990).
- Steiner, Richard C. 1991a. "Addenda to *The Case for Fricative-Laterals in Proto-Semitic*." In *Semitic Studies in Honor of Wolf Leslau*, ed. Alan S. Kaye, pp. 1499–1513. Wiesbaden: Harrassowitz.
- . 1991b. "The Aramaic Text in Demotic Script: The Liturgy of a New Year's Festival Imported from Bethel to Syene by Exiles from Rash." *Journal of the American Oriental Society* 111: 362–63 [bibliography in note 1].
- Strack, Hermann L. 1911. *Grammatik des Biblisch-Aramäischen*, 5th ed. Munich: Beck.

This page intentionally left blank.

Chapter 9

Classical Syriac Phonology

Peter T. Daniels
University of Chicago

9.1. Introduction

9.1.1. *The Syriac language*

The astute eighteenth-century Orientalist Johann David Michaelis recommended that those who would undertake the study of Semitic languages begin with Aramaic, because it is the simplest and most familiar—its syntax and vocabulary have been considerably influenced by Greek—but still embodies the main characteristics of the Semitic type; next proceed to Arabic, for it is the most elaborately developed yet also the most regular; and only then turn to Hebrew, the most important and attractive yet also the most difficult (1786: 21–26). This observation holds especially for the great literary language of the Aramaic group: Syriac.

Syriac is the liturgical language of many Eastern churches, founded on the dialect of Edessa (modern Urfa), seat of one of the first Christian kingdoms, Osrhoëne. Besides the standard Bible translation called the Peshitta and several earlier and later versions, an enormous devotional literature exists (preserved in manuscripts in monasteries throughout the Syrian desert, ranging from modern Turkey to Iran; Brock 1994), the creation of which was not interrupted by the Islamic Conquest of the 7th century C.E.; the last and perhaps greatest writer of Classical Syriac, Gregorius Bar Hebraeus, whose versatility rivaled that of Aristotle, lived in the 13th century.

There was a change, though, in the spoken language of the region following the Conquest: the Muslim majority spoke Arabic, with Aramaic-speaking Jewish and Christian minorities found in villages isolated in the mountainous terrain of Kurdistan. The written Aramaic languages continued in use in these communities, with relatively little influence from the spoken dialects. Moreover, as is often still observed in polycultural societies,

ACKNOWLEDGMENT: I am, as always in matters Aramaic, indebted to Robert D. Hoberman—who was my first teacher of Semitic linguistics, when we were fellow undergraduates.

there seems to have been little interaction among the Muslim, Christian, and Jewish populations and hence little convergence or mutual influence of their languages—so not much of an Arabic element can be found in Classical Syriac.

The literature in Syriac is periodized into a Golden Age, before the Conquest, and a Silver Age, afterward. The territory was partitioned between the Persian and Roman empires. During the Golden Age there came about a schism in the Syrian church, on Christological grounds, with the Persian (East) Syrians becoming Nestorian Christians and the Roman (West) Syrians Monophysite (or Jacobite) Christians. From the fifth century, these two communities had nothing to do with each other, and separate reading traditions, with distinct phonologies, grew up. The West Syrians maintained the Academy at Edessa and the East Syrians established their at Nisibis. The former seem to have maintained the historic forms more consistently (Nöldeke 1904: xxxi–xxxiv).

None of the Aramaic languages currently spoken seems to be a direct descendant of Syriac, though they use its script; but it remains the language of congregations as widely dispersed as the Martomite Christians of Madras, India (who take their name from the Apostle Thomas, whom legend names as the first missionary to the East), and the Aramaic-speaking diaspora of Flint, Michigan; Chicago, Illinois; Modesto, California; and, most recently, Western Europe, especially Sweden.

9.1.2. *The study of Syriac*

Our knowledge of the phonology of Syriac derives from four sources: from statements of Classical Syriac grammarians (Bar Hebraeus not least among them); from the notations in the texts prepared by the Syriac Masoretes; from the reading tradition as preserved by liturgical scholars; and from consideration of related languages and of loanwords to and from Syriac.

9.1.2.1. *Syriac grammarians*

The Syriac grammarians can be catalogued from the index entries in Baumstark 1922. Wright 1889 includes more information about the authors, but the generation between the two histories was notably productive in Syriac studies. Duval 1907, though less detailed than the other two histories, is arranged thematically rather than chronologically.

After some authors known only from mentions by later chroniclers, the earliest to write a grammar of which portions survive and have been pub-

lished is Jacob of Edessa (d. 708; ed. Phillips 1869, Wright 1871), representing the Jacobite tradition. (Mingana 1933: 251f. describes ms. 104—dated ca. 1840—as containing two works of Jacob; it seems not to be known to the earlier editors.) Richard Gottheil (1893) discovered fragments of a grammar by Davidh bar Paulos of Beth Rabban, from the late 8th century. Two Nestorians, contemporaries, are Elias of Tirhan (d. 1049; ed. Baethgen 1880) and Elias bar Shinaya (975–after 1049; ed. Gottheil 1887); and Joseph bar Malkon and Johannes bar Zo‘bi (both early 13th century) wrote major grammars that remain unpublished. Jacob of Tagrit (= Severus bar Shakko, d. 1241) composed a grammar (ed. Merx 1889) and a treatise on accentuation (ed. Martin 1879). Gregory Abu-l-Faraj, called Bar Hebraeus (1225/6–1286), composed a large prose grammar, the “Book of Rays” (latest ed. Moberg 1922 [apparatus in French], German trans. Moberg 1907–13; this is what is cited below), a shorter grammar in verse (ed. Bertheau 1843; Martin 1872, vol. 2), and at his death left unfinished a third, the “Book of Sparks.” The scholars most associated with study of the Syriac grammarians are the Abbé Paulin Martin (whose work tends to be deprecated by subsequent scholars, e.g. Nöldeke 1872), Adalbert Merx (his *Historia* is the only extended treatment), and J. B. Segal (his 1953 and 1989 studies of specific points of orthography consult both the manuscript tradition and the testimony of the grammarians). The Syriac grammatical tradition is dependent, first, on the Greek (specifically Dionysius Thrax, see Matthews 1994), and, later, on the Arab. A chapter on it—the first in any modern reference work—has been announced for the *Geschichte der Sprachwissenschaften* to be published by Walter de Gruyter; Kees Versteegh is the responsible editor.

9.1.2.2. *Liturgical tradition*

The received tradition among one of the churches that preserve Syriac liturgy has recently been studied by Hoberman (in press). He transcribes and analyzes the pronunciation of two expatriate Chaldean Catholic clergymen: tape recordings of a Chicago resident, and didactic materials prepared by a Californian. Mingana (1905) is regarded as a codification of the normative rules for the pronunciation of Syriac.

9.1.2.3. *Modern linguistic investigation*

Sokoloff (1978: 161) observed that the amount of work devoted to each of the Aramaic languages is in inverse proportion to the quantity of materials

preserved in each. While in recent years growing attention has been paid to contemporary spoken Aramaic (see Hoberman, this volume), it remains true that Syriac is the most neglected of the classical Semitic languages. The only full-scale grammar of modern times remains that of Duval (1881), which is exhaustive but betrays a premodern mindset. The standard reference is Nöldeke 1904 [1898], by the greatest of all scholars of Aramaic (and of much else Oriental). Crichton's translation includes some corrections and notes, but many fewer than his edition of Dillmann (1907). The marginalia in Nöldeke's copy of the book were transcribed (and lightly supplemented) by Schall (1966: 313–46) as an appendix to the reprint. The grammar is in fact, however, *Kurzgefaßte* (or *Compendious*—condensed, or abridged).

Study grammars of Syriac—textbooks—are in slightly fuller supply. The convenient *Porta Linguarum Orientalium* has included two avatars, Nestle (1889) and Brockelmann (1960; first edition, 1899). Nestle's has a more linguistic feel, and includes a very comprehensive bibliography both of the literature and of grammatical studies to date; Brockelmann's was reprinted as recently as 1976. The rival series *Clavis Linguarum Semiticarum* (which boasts the same general editor, Hermann L. Strack) includes Ungnad (1913), another manual informed by contemporary linguistics. All three of these contain reading selections and a glossary; Brockelmann's is made available to the English-speaking student by Goshen-Gottstein (1970).

The textbook most likely to be encountered by the student is Robinson 1962 (which follows the English model of lessons with translation exercises both out of and into Syriac but does not include a chrestomathy); it has been superseded by Healey (1986), using the inductive method. Kiraz (1994) includes a glossary of words occurring 10 or more times in the New Testament, verb paradigms, indexes, and a “skeleton grammar” by Sebastian P. Brock. Muraoka (1987) addresses the student whose teachers do not heed the advice of Michaelis. More than half of Eaton 1980 concerns Syriac studies; Brock (1980: 14–18) describes additional pedagogical materials.

There are two desk dictionaries, J. Payne Smith 1903 and Brockelmann 1928; the former is based on R. Payne Smith and Margoliouth 1879–1901 (supplemented, Margoliouth 1927). Brockelmann's etymological indications remain the fullest and most reliable available for Semitic. His and the large *Thesaurus*'s definitions are in Latin; J. Payne Smith's are in English.

An unwonted, welcome appearance of Syriac and other classical Aramaic data in the theoretical linguistic literature is found in Aronoff 1994.

9.2. Orthography

9.2.1. Consonants

The Syriac alphabet provides for the same 22 consonants as the Hebrew alphabet: <’ b g d h w z ḥ ṭ y k l m n s ‘ p š q r š t>. There are three different forms of the script (akin to the difference between, say, Roman type, the Fraktur formerly used for German, and the Irish alphabet): the older Estrangelo, the Serto of the Jacobites, and the Nestorian. They look different—a few of the letters even exhibit different basic forms—but the consonantal inventories are the same and orthographic principles do not vary (Daniels 1996).

9.2.2. Vowels

Syriac makes greater, more regular use of *matres lectionis* (vowel letters, i.e., consonant letters indicating vowels) than earlier forms of Aramaic. Virtually all long vowels other than *ā* are indicated within the consonantal text, as well as every occurrence of *o* and *u* (except in the two words *kul* ‘all’ and *meṭul* ‘because’).

There are in addition two different sets of optional vowel marks. The Nestorian script can be vocalized with a set of characters (developed from the diacritic point described in § 9.2.3) built from dots above and below the line of consonants: <a ā ē i o u>. The Serto script can be vocalized (“pointed,” the term used in Hebrew studies) with a set of characters derived from the Greek vowel letters: <a e i o u>. These are placed above or below the line of consonants, with specific positioning depending on esthetic considerations. (Modern publications tend to mix the styles of vowel and consonant, probably according to the fonts of type available to a particular printer.)

9.2.3. Diacritics

The Syriac script provides for diacritics on the orthographic, phonological, and morphological levels. Consonants that are not pronounced but are retained as part of historical orthography can be “canceled” with an over-score. Spirantized pronunciation of the six stops can be marked by a small dot under the letter, and plosive pronunciation by a dot above (Segal 1989).

Plural nouns are marked by a horizontal pair of dots above the word. Single dots above or below words otherwise spelled alike (but vocalized differently) distinguish various verb inflections. These two kinds of morphological markers tend to be used even in otherwise unpointed texts. The evolution of

the second kind from dots originally indicating broad vocalic categories (full/reduced, high/low) can be followed in the manuscripts (Segal 1953).

9.3. Segmental phonology

9.3.1. Consonants

The phonological change in Semitic that characterizes all (except the earliest) and only the Aramaic languages is the shift of the interdentalals to stops (in Hebrew and Akkadian they became sibilants, and in South Semitic they survived). The reconstructed segment *ʒ, whose reflexes vary in earlier Aramaic, is realized as ʕ in Syriac. [Examples hereinafter are generally taken from Nöldeke.]

The consonants shown in Table 9-1 are those notated in the Syriac script. The characterizing phrases in the list that follows are Bar Hebraeus’s descriptions. On the phonemic status of the fricatives, see § 9.4.1.

Table 9-1. Syriac Consonants

p b	t d	k g	ʔ
(f v)	θ ð	x ɣ	h
	ʈ	q	
	s z	š	ħ ʕ
	ʃ		
w		y	
	r		
	l		
m	n		

9.3.1.1. Labials

- b ≡ Voiced bilabial stop
- p ≡ Voiceless bilabial stop
- m ≡ Bilabial nasal

There was an additional voiceless labial, answering to Greek π, presumably completely unaspirated, discernible only in carefully pointed texts; in the absence of a symbol for an emphatic labial stop, ordinary ≡ was used (cf. ≡ for τ and ≡ for κ).

9.3.1.2. *Dentals*

- d ܕ Voiced dental stop
 t ܬ Voiceless dental stop
 ܬ Emphatic dental stop
 n ܢ Dental nasal

Bar Hebraeus (Tractate IV, chap. 1, sec. 3) describes these as apico-dental.

9.3.1.3. *Velars*

- g ܓ Voiced velar stop
 k ܚ Voiceless velar stop
 q ܩ Voiceless uvular stop

9.3.1.4. *Laryngeals*

- ? ܐ Glottal stop
 h ܗ Glottal fricative

Etymological (orthographic) ? is virtually never pronounced. Words spelled with initial ? are pronounced with initial vowel instead (#?C > #eC, #?i > #i, etc.) Intervocally, ? is [j].

9.3.1.5. *Pharyngeals*

- ʕ ܥ Voiced laryngeal fricative
 ܥ Voiceless laryngeal fricative

Bar Hebraeus does not distinguish the pharyngeals from the laryngeals in terms of place of articulation; he includes *h* with the fricatives and *h* with the sibilants as regards interruption of the airstream (loc.cit., sec. 4). Hoberman reports the pronunciation of Classical *h* and *ʕ* as [h] and [ʕ] by speakers of Colloquial Aramaic who use [x] and [ʔ] in corresponding vernacular words. He recognizes that “this fact ... suggests that these sounds existed continuously in the speech repertoire of the community even as they were changing to [ʔ], zero, or [x] in the vernacular (and were not merely reintroduced subsequently in borrowings from Arabic). Before the change the word for ‘life’ was [ħaːje] in both literary and colloquial registers. After the change it was [ˈxaːje] in the colloquial language but remained [ħaːje] in contexts that marked it as belonging to the literary register” (ms. p. 7, transcription adjusted).

9.3.1.6. *Emphatics*

Bar Hebraeus does not group the three emphatics *t*, *q*, and *s* in his lists of places or manners of articulation. According to Hoberman (pers. comm.),

they are realized as [tʰ, sʰ, q] in the modern Classical tradition and tend to affect neighboring segments, as in Arabic. Duval (1881: 42) and Nöldeke (1904: 39) suggest that the phenomenon of “flat” vs. “plain” words found in Modern Aramaic was operative or at least incipient in Syriac, but there is no orthographic evidence.

9.3.1.7. *Sibilants*

z	𐤆	Voiced apico-dental sibilant
s	𐤍	Voiceless apico-dental sibilant
ṣ	𐤌	Emphatic apico-dental sibilant
š	𐤒	Palatal sibilant

9.3.1.8. *Resonants*

l	𐤋	Lateral
r	𐤓	Apico-dental vibrant(?)

9.3.1.9. *Semivowels*

w	𐤅	Bilabial approximant
y	𐤄	Palatal approximant

Postvocally, *w* and *y* enter into diphthongs. Initial radical **w* shares in the Northwest Semitic change to *y*: *yeldaṭ* ‘she bore’, cf. Arabic *waladat*, but (causative) *awlaḡ* ‘he begot’. (The third person imperfect subject prefix is *n*- rather than **y*-, cf. *l*- in other varieties of Aramaic. This does not have a phonological explanation.)

9.3.2. *Vowels*

The distinction between East and West Syriac is seen in the respective vocalic systems. Hoberman (1992: 99) reconstructs an Early Syriac vowel system *ī ē ā ō ū ē a u* (*ē* merges with *ē* in the east and with *ī* in the west; see Blau 1969 on the “extra” *ē*). The East Syriac more closely reflects the inherited Aramaic pattern; the West Syriac seems to reflect an areal phenomenon that has persisted from ancient Canaanite through modern Arabic dialects: *ā > ō*.

9.3.2.1. *East Syriac vowels*

In its most elaborated form, the East Syriac vowel notation system (using combinations of dots) denotes the seven segments customarily transliterated as follows: *ă, ā, ě/ī, ē, ī, ŭ, ǝ*. As in transliterations of Masoretic Hebrew, the use of breves and macrons is misleading; not quantity, but quality is dis-

tinguished (Nöldeke 1904: 9). Likely phonetic equivalents would be respectively [a, ɒ, ε~ɪ, e, i, u/ʊ, o/ɔ]. Quantity is taken to depend on syllable structure.

9.3.2.2. *West Syriac vowels*

The West Syriac vowel notation system (using Greek letters) denotes the following five segments: *a*, *ō* < *ā*, *e*, *ī* (partly < *ē*), *u* (partly < *o*). Note the historical tendency to vowel raising: WS *nīmar*, ES *nēmar* ‘he says’, WS *šlūṭō*, ES *šlōṭā* ‘prayer’. This is the style of vocalization and pronunciation that first came to European attention, so the Syriac Bible in early works is called the “Peshitto.”

9.3.2.3. *Shwa*

Whether to claim the existence of shwa in Syriac depends on the phonological theory espoused. There are no symbols for notating shwa, so Bar Hebraeus is unable to describe it. But spirantization patterns (see immediately below) require the presence of some sort of vowel in many circumstances where none is written. For those who operate in terms of underlying representations, this may be a hypothetical vowel that is deleted in the process of deriving surface forms; for those—including Nöldeke—who wish to interpret Syriac as though it were Biblical Hebrew (or, at least, Biblical Aramaic), there is much talk of “vocal shwa” or “*shwa mobile*”; and for those who place Syriac in the context of the entire history of Aramaic, the evidence of the orthography is taken literally, and Syriac is transliterated without shwa (so Hoberman, after Kaufman 1984, followed by Daniels 1996). Evidence for the absence of shwa is voicing assimilation affecting consonants originally separated by a shwa (Nöldeke 1904: 15). In Hoberman’s corpus, only in a sequence of three or more consonants is an epenthetic shwa inserted, and that before the next to last consonant (without regard for etymology): [umma:rəṭʰo:n] ‘and he said to them’, = *wmar lhon*.

9.4. Phonological processes

9.4.1. *Spirantization*

Characteristic of all of Aramaic (and probably borrowed thence into Hebrew) is the fricativization, or lenition, of any non-lengthened non-emphatic postvocalic stop, a process traditionally called spirantization and usually transliterated either with a barred letter (*ḅ ḡ ḏ*) or with *h* (*bh gh dh*). Because

it is predictable from syllabic structure, one might suppose that spirantization need not be noticed in transcription or in linguistic work—except for the fact that it constitutes the only distinction between the first person singular and the second person masculine singular in the perfect tense of so-called “third weak” verbs: *rmīt* ‘I threw’ vs. *rmīt* ‘thou (m.) threwest’. (Further examples, with a historical account and noting of exceptions, may be found in Nöldeke 1904 § 23.) Spirantization is thus “phonemic,” and so long as it is consistently marked, it serves to indicate the underlying, or historical, vocalic patterns of Syriac words, and shwa need not be marked.

The spirantized version of *b* is not [v], but [w]. East Syriac *p̄* is given a spirantized pronunciation in only a handful of words; and in those few words, it is not [f] as might be expected, but [w]. (Hoberman states that Maronite Aramaic uses [b] for *b* and *b̄*, and [f] for *p* and *p̄*, presumably as in Arabic.)

9.4.2. Assimilation

Consonant clusters tend to exhibit regressive assimilation: *nezke* [neske] ‘conquers’, *dazkaryā* [dasxarja:] ‘of Zacharias’, *hesdā* [hezda:] ‘disgrace’; even *š* [ʒ] as in *hušbānā* [huʒba:na:] ‘an account’.

The *t* of the reflexive(-passive) prefix *et-* assimilates in emphasis or voice to a following dental first radical (cf. *etq̄tal* ‘was killed’): *eṭṭaššē* ‘was concealed’, *neddakrāk* ‘remembers thee’, but *ettkar* ‘remembered’.

Similarly, a final dental stop radical coalesces with *t* of the feminine suffix *-tā*: *pšītītā* [pʃi:ta:] ‘simple’, *ḥadtā* [ħaθa:] ‘new’.

Nearly every preconsonantal *n* assimilates to the following consonant.

Final radical *ʕ*, *h*, *ḥ*, and *ʔ*, when closing a syllable, change preceding *e* to *a*: *dabbah* ‘sacrificed’, cf. *qaṭṭel* ‘murdered’.

9.4.3. Metathesis

The *t* of the aforementioned reflexive prefix metathesizes with an initial radical sibilant: *estḥar* ‘was thought’ (cf. *šḥar* ‘thought’), *eštḥī* ‘was taken prisoner’, *eṣṭleḥ* ‘was crucified’, *ezdakkī* ‘was justified’.

9.4.4. Dissimilation

Within a root, *ʕ* followed by another *ʕ* becomes *ʔ* and is treated accordingly (this normally arose when one or the other *ʕ* < **ʕ*), as in (*ʔ*)*elʕā* ‘rib’ < *ʕelʕā* (so Jewish Aramaic; cf. Hebrew *šelʕ*, Arabic *ḍil(a)ʕ*). In West Syriac only, *ʕ* dissimilated to *ʔ* before *h* as well, as is sometimes indicated by spelling variants like *ḥyr* for *ʕhr* ‘to be in heat’.

9.5. Stress

Nöldeke reports (1904: 40) that among Nestorians, stress is always on the penult; but among the Maronites, stress is on a final closed syllable, otherwise on the penult. Hoberman reports the latter pattern as normative today among Chaldeans, noting that the final vowel is short unless stressed—i.e., the only final long vowels are in monosyllables, e.g. [ˈla:] ‘not’, [ˈbi:] ‘in me’.

9.6. Intonation

One might not expect to be able to recover information as to intonation patterns in a language that has not been in current use for nearly a millennium and a third, but fully pointed Syriac texts are provided with a series of “accents” that are intended to resolve syntactic ambiguities (Bar Hebraeus’s introductory passage is translated by Segal 1953: 61–62); and from Bar Hebraeus’s detailed account (loc.cit. chap. 6), it just *may* be possible to extract some information about the intonation of the affected clauses.

References

- Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional Classes* (Linguistic Inquiry Monograph 22). Cambridge: MIT Press.
- Baethgen, Friedrich. 1880. *ܐܪܡܝܐ ܕܡܪ ܝܗܝܠܐ ܐܝܬܝܗ ܐܘܪܝܬܐ ܕܡܪ ܝܗܝܠܐ ܐܝܬܝܗ* oder syrische Grammatik des Mar Elias von Tirhan. Leipzig: Hinrichs.
- Baumstark, Anton. 1922. *Geschichte der syrischen Literatur*. Bonn: Marcus & Weber.
- Bertheau, Ernestus. 1843. *Gregorii bar Hebraei qui et Abulpharag grammatica linguae syriacae in metro Ephraemo*. Göttingen: Vandenhoeck & Ruprecht.
- Blau, Joshua. 1969. “The Origins of Open and Closed e in Proto-Syriac.” *Bulletin of the School of Oriental and African Studies* 32: 1–9.
- Brock, Sebastian. 1980. “An Introduction to Syriac Studies.” In Eaton 1980: viii, 1–33.
- . 1994. “The Development of Syriac Studies.” In *The Edward Hincks Bicentenary Lectures*, ed. Kevin J. Cathcart, pp. 94–113. Dublin: University College, Department of Near Eastern Studies.
- Brockelmann, Carl. 1928. *Lexicon syriacum*, 2nd ed. Halle: Niemeyer. Repr. Hildesheim: Olms, 1966.

- . 1960. *Syrische Grammatik mit Paradigmen, Literatur, Chrestomathie und Glossar*, 12th ed. Repr. Leipzig: VEB Verlag Enzyklopädie, 1976.
- Daniels, Peter T. 1996. "Aramaic Scripts for Aramaic Languages: Classical Syriac," in *The World's Writing Systems*, ed. Peter T. Daniels and William Bright, pp. 499–504. New York: Oxford University Press.
- Dillmann, August. 1907. *Ethiopic Grammar*, 2nd ed., enlarged and improved, ed. Carl Bezold, trans., with additions, James A. Crichton. London: Williams and Norgate. Repr. Amsterdam: Philo, 1974.
- Duval, Rubens. 1881. *Traité de grammaire syriaque*. Paris: Vieweg.
- . 1907. *La littérature syriaque des origines jusqu'à la fin de cette littérature après la conquête par les arabes au XIIIe siècle*, 3rd ed. Paris. Repr. Amsterdam: Philo, 1970.
- Eaton, John, ed. 1980. *Horizons in Semitic Studies: Articles for the Student* (University Semitics Study Aids 8). Birmingham, England: University of Birmingham, Department of Theology.
- Goshen-Gottstein, Moshe. 1970. *A Syriac-English Glossary with Etymological Notes Based on Brockelmann's Syriac Chrestomathy*. Wiesbaden: Harrassowitz.
- Gottheil, Richard J. H. 1887. *A Treatise on Syriac Grammar by Mâr(i) Eliâ of Šôbbâ*. Leipzig: Drugulin.
- . 1893. "Dawidh bar Paulos, a Syriac Grammarian." *Journal of the American Oriental Society* 15 (1893) cxi–cxviii.
- Healey, John F. 1986. *First Studies in Syriac*, corrected repr. ([Birmingham] University Semitics Study Aids 6). Sheffield: Sheffield Academic Press.
- Hoberman, Robert D. 1992. "Aramaic." In *International Encyclopedia of Linguistics*, ed. William Bright, pp. 98–102. New York: Oxford University Press.
- . in press. "The Modern Chaldean Pronunciation of Classical Syriac." In *Humanism, Culture, and Language in the Near East: Studies in Honor of Georg Krotkoff*, ed. Asma Asfaruddin and A. H. Mathias Zahniser. Winona Lake, Ind.: Eisenbrauns.
- Kaufman, Stephen A. 1984. "On Vowel Reduction in Aramaic." *Journal of the American Oriental Society* 104: 87–95.
- Kiraz, George Anton. 1994. *Lexical Tools to the Syriac New Testament* (*Journal for the Study of the Old Testament Manual* 7). Sheffield: Sheffield Academic Press.

- Margoliouth, J[essie] P[ayne Smith]. 1927. *Supplement to the Thesaurus of R. Payne Smith*. Oxford: Clarendon.
- Martin, Abbé [Paulin]. 1872. *Oeuvres grammaticales d'Abou 'Ifaradj dit Bar Hebreus*. 2 vols. Paris: Maisonneuve.
- . 1879. *De la métrique chez les Syriens* (Abhandlungen für die Kunde des Morgenlandes 7/2). Leipzig: Brockhaus.
- Matthews, P. H. 1994. "Greek and Latin Linguistics." In *History of Linguistics*, ed. Giulio Lepschy, vol. 2: *Classical and Medieval Linguistics*, pp. 1–133. London: Longmans.
- Merx, Adalbert. 1889. *Historia artis grammaticae apud Syros* (Abhandlungen für die Kunde des Morgenlandes 9/2). Leipzig. Repr. Nendeln, Liechtenstein: Kraus, 1966.
- Michaelis, Johann David. 1786. *Abhandlung von der syrischen Sprache, und ihrem Gebrauch*, 2nd ed. Göttingen: Vandenhock.
- Mingana, Alphonse. 1905. *Clef de la langue araméenne ou grammaire complète et pratique des deux dialectes syriaques occidental et oriental*. Mosul: Imprimerie des Pères Dominicains.
- . 1933. *Catalogue of the Mingana Collection of Manuscripts Now in the Possession of the Trustees of the Woodbrooke Settlement, Selly Oak, Birmingham*, vol. 1: *Syriac and Garshūni Manuscripts*. Cambridge: Heffer.
- Moberg, Axel. 1907–13. *Buch der Strahlen: Die grössere Grammatik des Barhebraeus*. Leipzig: Harrassowitz [Einleitung und Zweiter Teil, 1907; Erster Teil und Stellenregister, 1913].
- . 1922. *Le livre des splendeurs: La grande grammaire de Grégoire barhebraeus* (Acta Reg. Societatis Humaniorum Litterarum Lundensis 4). Lund: Gleerup.
- Muraoka, T. 1987. *Classical Syriac for Hebraists*. Wiesbaden: Harrassowitz.
- Nestle, Eberhard. 1889. *Syriac Grammar with Bibliography, Chrestomathy and Glossary*, 2nd ed., trans. Archibald R. S. Kennedy (Porta Linguarum Orientalium 5). Berlin: Reuther.
- Nöldeke, Theodor. 1872. [Review of Martin 1872]. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 26: 828–35.
- . 1904. *Compendious Syriac Grammar*. Translated by James Crichton. London: Williams & Norgate.
- Payne Smith (Mrs. Margoliouth), Jessie. 1903. *A Compendious Syriac Dictionary*. Oxford: Clarendon.

- Payne Smith, R., and D. S. Margoliouth. 1879–1901. *Thesaurus syriacus*. Oxford: Clarendon.
- Phillips, George. 1869. *A Letter of Mār Jacob, Bishop of Edessa, on Syriac Orthography* London: Williams and Norgate.
- Robinson, Theodore H. 1962. *Paradigms and Exercises in Syriac Grammar*, 4th ed., rev. L. H. Brockington. Oxford: Clarendon.
- Schall, Anton, ed. 1966. “Die handschriftlichen Ergänzungen in dem Handexemplar Theodor Nöldekes und Register der Belegstellen.” In Theodor Nöldeke, *Kurzgefaßte syrische Grammatik*, 2nd ed. (1898), repr. Darmstadt: Wissenschaftliche Buchgesellschaft, pp. 307–401.
- Segal, J. B. 1953. *The Diacritical Point and the Accents in Syriac* (School of Oriental and African Studies, London Oriental Series 2). London: Oxford University Press.
- 1989. “*Quššaya* and *Rukkāka*: A Historical Introduction.” *Journal of Semitic Studies* 34: 483–91.
- Sokoloff, Michael. 1978. “The Current State of Research on Galilean Aramaic.” *Journal of Near Eastern Studies* 37: 161–67.
- Ungnad, Arthur. 1913. *Syrische Grammatik mit Übungsbuch* (Clavis Linguarum Semiticarum 7). Munich: Beck.
- Wright, William. 1871. *Fragments of the Syriac Grammar ... of Jacob of Edessa*. London(?): privately printed in an edition of 50 copies; reproduced [apparently not in facsimile] in Merx 1889: 𐤀𐤁𐤁𐤃–𐤀𐤁𐤁𐤄 [73–84].
- . 1889. “Syriac Literature.” *Encyclopædia Britannica*, 9th ed., 22: 824–56. Repr. separatim.

Modern and Classical Mandaic Phonology

Joseph L. Malone

Barnard College and Columbia University

10.1. Introduction

This chapter deals with two types of Mandaic, a Northwest Semitic language of the Eastern Aramaic subfamily: Modern Mandaic (MM), still spoken residually in southern Iran and not long ago in southern Iraq as well; and Classical Mandaic (CM), which flourished over roughly the same terrain for several centuries up to the Islamic ascendancy, leaving a rich heritage of sacred texts still used by the Mandaean religious community. An intermediate form of the language, Postclassical Mandaic, will be occasionally mentioned but not treated in its own right.

The following pages depend crucially on the work of Rudolf Macuch: Drower & Macuch 1963 (D), Macuch 1965 (H) (of which the *Vocabulary of the Vernacular*, pp. 489–526, will be tagged as VV), and Macuch 1989 (C).

(This may be a suitable spot for listing some other abbreviations to be employed throughout: # = word or clitic boundary; C = consonant (when not tagging Macuch 1989); G = ‘ghost consonant’ (§ 10.4); $\check{V}$, $\dot{V}$, $\bar{V}$ = reduced (§ 10.2.2), short, long vowel respectively; V = V of any length; $\acute{V}$ = stressed V (most often not indicated); $\sqrt{\quad}$ = root, radical (Semitic internal-flective, non-concatenative lexeme or segment thereof); $\langle X \rangle$, $t[Y] = X$ is orthographic, Y is a traditional pronunciation of CM (§ 10.3.0); in rules iA , jB , kC , ... = rule i precedes j precedes k ..., as a special case of which B^* marks a ‘circuit’ rule (§ 10.3.2); $X \neg Y = X$ precedes Y (§ 10.2.4).)

MM will be treated in § 10.2, CM in § 10.3. A few theoretical matters will be briefly touched on in § 10.4 as an appendix.

Table 1 is a (broad-)phonetic chart for MM and CM segments compositely, parenthesized segments being exclusively MM.

Table 10-1. Mandaic Segments

p	t	ṭ ^a	(č)	k	q ^a
b	d	(ḏ)	(j)	g	
	f	θ	s ʃ š	x	(ḥ)
β	ð ^b	z	(z) (ž)	ɣ	(‘)
m	n			ŋ ^c	
	l				
	r				
w		y			
		i ^d		u	
		e ě (or ě)	ō	o	
		æ ^e	ă	ɔ	
			a		

^a‘Emphatics’; analyzed as uvularized in Malone 1976.

^bMM only in Arabic borrowings; otherwise exclusively CM (§ 10.2.3).

^cPrevelar allophone of /n/.

^dVowels other than *ĕ* may appear as short or long.

^e[æ, a, ɔ] appear to be mutual allophones, though [ɔ] and [æ] sometimes also vary with [o] and [e] respectively.

10.2. Modern Mandaic

10.2.1. Consonant groups

While MM is quite tolerant of two-member clusters in word-final position, it is quite stingy with the same word-initially. Of the 3000-odd entries in VV, this was found in only six lexemes (*šβīr* ‘good’, *trēn* ‘two’, *smāla* ‘left’, *sfargšlā* ‘quince’, *drafsā* ‘cross’, *šboroxta* ‘life’) and twice in the morphosyntactic combination durative proclitic *q-* plus participle (*qšāyel* ‘spin’, *qyādī* ‘know’).

It is not clear what generalizations might be made here. For one thing, some of the items show alternatives to #CC-: in particular #CVC- (*ššβīr*, *qšmāyeš* ‘suck’), #CVC- (*qabāyeš* ‘stay’ [s.v. ‘silent’]), and #VCC- (*ešbīr* [see § 10.2.3 for the *b*], *esmāla*). Moreover, these alternations show up in other morphemes involving the same Cs (e.g. *dšrā* ~ *derā* ‘lift’ (G:213), cf. *drafsā*).

When intervocalic in clearly word-internal position, CC is unconstrained. However, the question arises as to what bearing proclitics might have on the pertinent (prosodic) notion of ‘word’ involved here. Provisionally, it seems as if a proclitic forms a bond with its host word looser than that contracted between clearly word-internal morphemes, but not so loose as that across independent words. Thus though the proclitic *la-*, *læ-* ‘not’ most frequently (though not invariably (C:216)) combines with a past tense verb in its #CVC- or #CVC- form (*læ-dehelt* ‘you didn’t fear’ (C:211), rather than *læ-dhelt*), the #CC- form appears in the durative (*la-qdahlā* ‘she doesn’t fear’ vs. *qādhel* ‘he fears’ (for the *a* ~ *ā*, see § 10.2.5)). (The *q-* ~ *qā-* are themselves proclitic, by the way, as in *qšāyel* and *qyādī* seen earlier. For more on this, see § 10.2.2 rule 1A.)

Even as there are thus organic relations among #CC-, #CVC-, #CVC- (occasionally #CVC-), and #VCC- (occasionally #VCC-) initially, so we find analogous relations among at least some of -CCC-, -CCVC-, -CCVC- (occasionally -CCVC-), and -CVC- medially. To these we turn next (§ 10.2.2). Here finally be it only noted that VV contains just one instance of -CCC-, *ginztær* s.v. ‘utmost’, analytically ‘much’ plus comparative suffix *-tær* borrowed from Persian. (Perhaps this functions as an enclitic in MM, in which case *nz* might be effectively acting as (quasi) word-final.)

10.2.2. Rules for consonant groups

Pending further research, alternations of the type just considered (excluding *Ṽ* ~ *Ṽ* components, for which see § 10.2.5) will be tied together by three rules: **Reduction**, which either weakens a vowel or deletes it altogether; **Promotion**, which strengthens a *Ṽ* to full-vowel status; and **Prothesis**, which resyllabifies the initial portion of a word through introduction of an anlaut vowel. Reduction must be able to precede the other two rules, since a Reduction-introduced *Ṽ* is subject to Promotion or to loss as a byproduct of Prothesis.

- (1A) **Reduction.** Most clearly in verbs (including participles) and proclitics, a non-long normally open-syllabic vowel is either deleted or weakened to *Ṽ*.

The alternatives in 1A are largely dictated by syllabic canons, deletion being mandatory when weakening would result in the impermissible sequence VCVCV, and weakening being mandated when deletion would entail the sequence CCC (unattested except for *ginztær* (§ 10.2.1), which however does not involve Reduction). On the one hand compare *yehem* ‘he sat’ (C:225),

ahaβ ‘he gave’ (H:550) with their feminine counterparts, where the second schematic (stem vowel) has been deletion-Reduced upon suffixation with *-at*: *yehmat* ‘she sat’ (C:225), *ahβat* ‘she gave’ (H:550). On the other hand compare various cases of proclisis of durative *qa-*, whose *a* remains in *qam-bašqer* ‘he recognizes’ (C:207) unweakened (as normally, its syllable being closed), or in fact does weaken (a rarity except perhaps in the proclitic /*qa-/*, as here) as per *qšmbašqernānī* ‘we recognize him’ (C:207), but in neither case deletes, which would entail impermissible *qmb*—and then also, as seen in § 10.2.1, either weakens in *qšmāyeš* or deletes in *qšāyel*, again however neither option entailing CCC. (In this vein, another symptom of the full-word vs. part-of-word ambivalence of proclitics discussed in § 10.2.1 may be seen in the occasional failure of open-syllabic *qa-* to reduce; e.g. *qabayeš* ‘stays’ (H:519). Such retention may be found even prevocally (over a ‘ghost’ consonant (G), see appendix (§ 10.4)), e.g. *læ-eβādnī* ~ reduced *la-βādnī* ‘we didn’t do it’ (C:47), from /*la-Gšβadnī/*.)

Though Reduction is formulated to introduce *ǃ*, this does not preclude there being underived *ǃ*s as well. In fact, assuming such *ǃ*s whenever it can be gotten away with is one way of cutting down on arbitrary analytic choices for phonological representations. And as a case in point, we may assume that many forms whose first vowel is phonetically *ǃ*, like most of those seen in § 10.2.1, show the same *ǃ* phonologically: thus /*ššβīr/*, /*dšrā/* (or perhaps /*dšray/*, cf. Malone 1985).

(2D) **Promotion.** Under various circumstances, a *ǃ* advances to full-vowel status, especially in a medial or closed syllable.

Before looking at some examples, a few words about *ǃ* are in order. In MM as in the classical Aramaic languages, this refers to what in German is called ‘Murmelvokal’, a (somewhat) mided and centralized vowel shorter in duration and laxer than its full-vowel counterparts. *ǃ*s may be ‘colored’ or ‘bleached’, the latter seeming to be the unmarked case in MM, surfacing as [ǃ]. While in some languages (e.g. Tiberian Hebrew (Malone 1993)) full vowels reduced to *ǃ* may retain distinctive coloring, in MM they seem not to—though they do sometimes pick up new coloring from enviroining consonant positions. In fact in C (though not in H) Macuch provides special symbols for these: ^a, ^o in this paper retranscribed as *ǃ̃*, *ǃ̄*. Since underlying *ǃ*s are assumed to be /*ǃ̄/*, *ǃ̃* and *ǃ̄* are derived by rule:

(3B) **Schwa Coloring.** *ǃ̃* may labialize (*ǃ̣̃*) or lower (*ǃ̤̃*) under the influence of an adjacent C position.

This much said, consider the derivation of three of the four allomorphs of the word for ‘white’, underlyingly /həwārā/: (i) the null case, = [həwārā]; (ii) by Schwa Coloring (due to w), → [hōwārā]; (iii) by Schwa Coloring followed by Promotion, → hōwārā → [howārā].

The fourth possibility is Promotion **without** Schwa Coloring, which gives [hewārā]—the default color of ə being closest to unrounded front mid (cf. Chomsky & Halle 1968, Malone 1993). (All forms from C:216.)

While the ‘circumstances’ mentioned in 2D remain to be discovered and systematized, it seems clear that Promotion of ǃ is the unmarked choice—most notably in medial syllables. (The only forms I have found to surface with medial ǃ are *sfargəḏlā* (§ 10.2.1), *yardəṇā* ‘Jordan’ (C:49), *mazgəḏdā* ‘mosque’ (H:511), *aβγəṣī* ‘stop (fem.)!’ (C:204), and *ahləxon* ‘your (pl.) family’ (C:195).)

Some further examples. From § 10.2.1: /dǃrā/ either = [dǃrā] or 2D → [derā]. From under 1A: /yǃhem/ 2D → [yehem], /yǃhemat/ 1A → yǃhmat 2D → [yehmat] (Promotion is virtually obligatory in closed syllables; exceptions like [qǃmbašqernanī] are rare, usually involving proclitic qǃ-; if 2D did apply here, [qe-] would be predicted, as in fact in [qe-mbāsquerettellī] ‘you (pl.) know him’ (C:182); thus [qa-] in homoparadigmatic [qambašqer] is not via 2D, but rather retains phonological /qa-/); /Gǃhaβ/ 3B → Gǃhaβ 2D → [ahaβ], /Gǃhaβat/ 1A → Gǃhβat 3B → Gǃhβat 2D → [ahβat] (the shadow consonant G (§ 10.4) frequently triggers ǃ-coloring 3B for historical reasons, being the reflex of a lost *ʔ (as here) or *ʕ). From H(281–82): /qamǃhambelet/ 1A → qamǃhambǃlet 2C → [qamhambelet] ‘you profane’: /qamǃhambalat/ 1A → qamǃhambǃlet 2C → [qamhambelet] ‘you are profaned’ (note the neutralization of the opposition between these forms, still maintained in the masculine [qamhambel] vs. [qamhambal] (loc. cit.); note also the deletional Reduction (1A) of /ǃ/, relieving an impermissible sequence VCǃCV).

(4C) **Prothesis**. An initial group #C(ǃ)C may resyllabify to #ǃCC.

Though Prothesis itself is relatively rare, it appears to categorically feed Promotion (2D). Examples: /hǃroβ/ may undergo 3B → hǃroβ or not = hǃroβ, whereupon hǃroβ may undergo 4C → ǃhroβ and all three alternatives undergo 2D → [heroβ], [ahroβ], [haroβ] ‘he was spoiled’ (H:260); /nǃheθ/ –3B±4C → ǃnheθ ~ nǃheθ 2D → [enheθ] ~ [neheθ] ‘he descended’ (H:499).

The formulation in 4C states ‘initial’, rather than specifically ‘word-initial’, thus hedging on whether the rule may apply following a proclitic, cases like [qa-erqīha] analytically ‘in heaven’ (H:505) for default of alternations

being ambiguous as to whether ← /erqīhā/, /rəqīhā/, or even /rqiḥā/ (for the $a \sim \bar{a}$, see § 10.2.5). (In vein of the last possibility, note the Prothesized variant [esmāla] to [smāla] in § 10.2.1. Should the underlying representation of such #CC forms be as per null-case /smālā/ or rather /səməlā/ with lexically governed deletion-Reduction (1A)? The question must be left open here.)

Mention should also be made of a special process for alleviating -CC(Ṽ)C- when the first two Cs constitute a geminate, a process which moreover extends to dispelling geminate -CC# in word-final position:

- (5B) **Simplification.** A geminate CC simplifies to C when in word-final position (___#) or when preceding a consonant either immediately (___C) or mediately over a reduced vowel (___ṼC). In the latter case, moreover, the reduced vowel is elided.

Examples: /barrexat/ 1A → *barrəḫat* 5B → [barxat] ‘she blessed’, cf. /barrex/ = [barrex] ‘he blessed’ (both H:264); /rabb/ 5B → [rab] ‘large’ (predicative form), cf. /rabbā/ = [rabbā] ‘large’ (attributive form) (both H:490). (For the inertness of $\dot{\text{a}}$ in 4C, 5B, and various other rules below, see § 10.4.)

10.2.3. *Spirantization*

Very little is left in MM of the venerable classical Aramaic process of postvocalic spirantization of simple non-emphatic (possibly unuvularized (Malone 1976)) stops. Putting aside for the moment certain innovations to be picked up later ($q \sim x$, $t \sim h$), the only de facto alternations I have found are these (in most cases citations are not exhaustive):

(6)		Stops	Spirants
a.	√zbn ‘transact’	<i>zabben</i> (C:221)	<i>zəḫban</i> (C:221)
	√bGy ‘want’	<i>qa-bā</i> (C:160)	<i>qa-ḫā</i> (C:160)
b.	feminine gender nominal suffixes (H:206, C:59)	<i>-tā, -tī</i>	<i>-θā, -θa, -θī;</i> <i>-aθ, -θe;</i> <i>-āθī, -āθ, -āθa</i>
c.	√sbr ‘good, beautiful’	<i>ešbīr</i> (H:504) <i>šbīrtær</i> (H:498) <i>ešbertī</i> (H:493)	<i>š(ə)ḫīr</i> (H:493, 504) <i>ešḫertā</i> (H:511)
d.	√byṣ ‘stop’	<i>beyeṣ</i> (C:204) <i>abyeṣ</i> (H:521)	<i>aḫyṣī</i> (C:204)
e.	√rb̄b ‘great’	<i>rab</i> (H:490) <i>rabbā</i> (H:490)	<i>raḫī</i> (H:3)

Clearly there are traces here of the old rule, which was by hypothesis still alive in CM (see § 10.3.3). Thus in 6a the pair *zabben* ‘sold’ and *zəβan* ‘bought’ is unchanged from CM, where only β of the latter reflects spirantization of a *b* which is both postvocalic and simple (non-geminate). The alternation in *qa-bā ~ qa-βā* might likewise be tractable when the ambivalence of proclitics is taken into account (§ 10.2.1). Moreover in 6b the constancy of spirantal θ postvocally in the singular construct (*-aθ*) and the plural (*-āθī*, *-āθ*, *-āθa*) also works out identically to the situation in CM. However, the overall disposition of the singular non-construct (*-tā*, *-tī*; *-θā*, *-θa*, *-θī*) can no longer be captured by a fully well-behaved phonological rule, despite the clearly persisting tendency for θ to occur postvocally and *t* postconsonantly. Note e.g. *-ta* in *baratta* ‘daughter’ (C:204) but *-θa* in *hadaθθa* ‘story’ (C:214) (and *-ta* again in the VV-given variant of the latter, *hadaθta*). Note also post-consonantal *-θa* in *demehθa* ‘tear’ (VV). A plausible way of responding to this situation in MM is with a **lexical redundancy condition** along the following lines:

- (6) The feminine nominal suffix unmarkedly subcategorizes for either of two shapes, /t/ postconsonantly and /θ/ postvocally.

Since 7 is lexical, it may be overridden by contrary lexical stipulations; e.g., /demeh/ would subcategorize for /θa/. Similarly the alternations in the other sets of 6 are, with one exception, best interpreted as synchronically lexical—though in part fossilized residue of the once living rule (notably in 6a) and in part governed by another redundancy condition:

- (6) In non-rilled non-uvularized obstruent internal-flective root consonants other than voiced apical, 1√C is unmarkedly occlusive while 2√C and 3√C are normally spirantal.

This condition is worded to exempt rilled $\check{c}$, *j*, *s*, $\check{s}$, *š*, *z*, $\check{z}$, *z*, and uvularized *t*, *q*, as well as *d* which in MM is reflex of the merger < **d*, **ḏ*. (One other change in phonetic content, incipient * θ > *h*, clings to a tenuous synchronic existence in the dialect variation *yitem ~ yehem* ‘sat’ (C:41).) While 8 does not hold for 6c (where it would predict only β) nor the β of 6d, it does hold for the β of 6a as well as for *b* and *y* in 6d. The *bb* in 6a, *e* might possibly reflect a third, overriding condition favoring occlusion of geminates (as in CM), but this is by no means certain since MM sports quite a few spirantal geminates, especially $\theta\theta$; e.g. *hadaθθa* (seen earlier), *maθθed-nax* ‘I’ll lift you’ (C:200), *aθθī* ‘brought’ (C:200), *eθθa* ‘woman’ (C:245);

but also *ff* in *affeq* ‘brought out’ (C:239) (for *ββ* and *xx*, see 10A, 11B). The *b* in *rab* of 6e would follow either from the possible condition favoring occlusive geminates (*rab* deriving from /rabb/ by 5B), or from lexical stipulation preempting 8.

This leaves only *raftī* in 6e uncovered. Following Macuch’s lead in H:38, 56–57 I think we may be dealing with a rule here. Not knowing the precise conditions, I make bold to formulate broadly:

- (7D) **Cluster Balancing.** A heterophonous cluster XY, X a nonrilled obstruent (especially *b*, *β*, *q*) and Y a stop (especially *t*), may dissimilate X in manner to become a spirant but assimilate X in voice.

Thus in 6e, /rabbtī/ 5B → *rabtī* → [raftī]; also *haftella* → [haftella] ‘you gave it’ (C:38); *bedaqtī* → [bedaxtī] ‘I put it’ (C:42). (It is possible that the spirantization and devoicing in 9D are independent processes, that the latter may be fed by the former, and be variably conditioned by other environments than /___C: cf. [genaβ] ~ [genaf] ‘steal’ (VV) vs. invariant [rab] ≠ non-occurrent [rap] ‘big’ (H:490); also √nsq in *nesxe-l-īd* ‘kissed the hand’ (C:42). On the other hand these may just reflect lexical quirks of √gnb and √nsq. Suggesting this may be occasional auslaut *t* in lieu of expected *d* (e.g. (C:193).)

Two other rules involving spirants in MM:

- (8A) **Reciprocal Assimilation.** In a group *abb*, *a* may regressively labialize to *o* and *bb* progressively open (spirantize) to *ββ*.

E.g. *tabbar* → [toββar] ‘break’ (VV), *habbet* → [hoββet] ‘mix’ (VV). (Note that [zabben] (above) is an exception.)

- (9E) **Spirant Assimilation.** *hx* may assimilate to *xx*, and *dθ* to *θθ*.

The sole examples involve the roots √ghx ‘laugh’ and √hdθ ‘tell’, but it may be that only these roots provide the sequences *hx* and *dθ* in MM: *gehxt* → [gexxat] ‘she laughed’ (C:208); *qa-mhaddθā* 5B → *qa-mhadθā* → [qa-mhaθθā] ‘(I (fem.)) tell’ (C:215).

It may be noted in passing that 10A and 11B add [ββ] and [xx] to [θθ] and [ff] (see above) as permissibly geminate slit spirants in MM. Recalling that [ð] is absent from MM except in Arabic loans, this leaves unattested only [ɣɣ] (in the corpus available to me).

10.2.4. Stress

Aside from some Persian loans, which may retain their original ultimate accent (e.g. *pæšimān* 'sorry' H:139), my reading of Macuch (H:135–40, C:49) suggests that word stress in MM is for the most part given by this rule:

- (10E) **Stress Placement.** For a full vowel V, $V \rightarrow [+stress] / __X\#$, where X contains no more than one full V.

Thus from § 10.2.1: [šβ̥ir] (X contains only *r*), [dráfšā] (X contains *fšā*, whose only *ā* is a full vowel), [sfárgə̌lā] (X contains *rgə̌lā*, where again only *ā* is a full vowel), [šboróxta], [ésb̥ir], [dérā] ~ [dǽrá] (X is empty and *ǽ* is unstressable because not a full vowel).

Notice that [ésbīr] and [dérā] show that Stress Placement must follow both Prothesis (4C) and arguably normally Promotion (2D). The qualification ‘arguably normally’ is called for in the case of forms like *aθā́*, *aθā́n* ‘he, they came’ (C:49), which suggest that Stress Placement applies to such forms **before** the *a* has Promoted from *ǎ* (← *ǝ* by ghost consonant Schwa Coloring (3B)). If the totality of such forms should begin in vowels, it might be possible to continue the ordering 2D[−]12E and instead add to 12E the variable condition ‘V is not initial’ (variable in view of forms like [émar] ‘he said’ (H:135)). However the forms [šəbā́q] ‘left’ (C:138) and [benā́] ‘constructed’ (C:160) plus some remarks by Macuch at H:265 suggest that non-initial Promoted vowels may also sometimes remain unstressed. If so, the best response is perhaps to add to 12E as unmarked the following earlier ordering as a marked option:

- (11D) **Stress Placement.** (Same content as 12E, but marked in application)

10.2.5. Alternations of quantity

MM evinces alternations of both consonant and vowel length in part determined by the stress. One situation is rather simple to state:

- (12F) **Pretonic Shortening.** A geminate or long segment earlier in the word than the stress tends to shorten.

Thus [áθθī] ‘he brought’, but $a\theta\theta\acute{\imath}t\bar{u} \rightarrow [a\theta\acute{\imath}t\bar{u}]$ ‘she brought them’ (C:200); $\bar{a}\beta d\acute{e}tton \rightarrow [a\beta d\acute{e}tton]$ ‘you (pl.) do’ (C:194).

Rule 13F apart, the situation with vowels is as of this writing bewilderingly unclear to me (cf. also C:47), at least a few examples being found of / $\dot{V}$ / $\rightarrow$ [$\bar{V}$] and / $\bar{V}$ / $\rightarrow$ [$\dot{V}$] in just about all positions within the word. Covert

prosodic conditions may very well be at work here. However, two cases appear with sufficient frequency to warrant a provisional rule:

- (13F) **Length Adjustment.** In / $\acute{V}$ C(C)(V)/, either stressed / $\acute{V}$ / may lengthen and/or unstressed / $\bar{V}$ / may shorten.

Thus *át* = [át] ~ [át̃] ‘you’ (C:200), *ám̄r* = [ám̄r] ~ [ám̄r̃] ‘command’ (C:198), *eḇádyon* → [eḇádyon] ‘they did’ (C:193), *eḇádnī* = [eḇádnī] ~ [eḇádnī̃] ‘we did’ (C:193), *gáḇrā* [gáḇrā] ‘man’ (C:207), *árbin* = [árbin] ~ [árbiñ] ~ [árbin̄] ‘forty’ (C:199), *gabbārā* 13F → *gabārā* = [gabārā] ~ [gabāra] ‘hero’ (C:207).

10.2.6. *Alternations of vowel quality*

One pristine rule of Aramaic may still linger in MM, though rendered moribund by exceptions:

- (14A) **Assimilatory Lowering.** A short vowel may drop to *a* immediately preceding 3 $\acute{V}$ r (and by prediction 3 $\acute{V}$ h, though no examples could be found). That is, $\acute{V}$ becomes [+low] under regressive contact with a [+low] C, as long as the latter is lexically third radical.

Thus, two antithetical examples from the /CaCCeC/ (paēl, quadriconsonantal) verb: /tabber/ +15A → [tabbar] 10A[−]12E → [tóḇḇar] ‘break’ (VV), but /bašqer/ −15A[−]12E → [bášqer] ‘know’ (C:207). (This rule has arguably already disappeared as such in MM, living on merely as the lexical quirk of certain verbs. Of the twelve 3 $\acute{V}$ r verbs listed in the glossary of C (pp. 193–263), only one (√kmr) is consistently +15A, another (√tḇr, above) is mixed, yet another (√dxr) is moot, and the balance are arguably or certainly −15A.)

Beyond 15A, MM contains a striking amount of variation in vowel color, especially in the verb, where moreover the variation is notably compounded by interaction with Length Adjustment (14F) and allied processes. Thus after Promotion (2D) from *ḡñon*, we have [genōn] ~ [genon] ~ [genōn̄] ~ [ginōn̄] ~ [ḡīnān̄] ‘they slept’ (C:210). (The ‘allied process’ responsible for $\acute{V}$ ~ $\bar{V}$ in the suffixal vowel here is provisionally lexical allomorphic.)

Whatever processes may ultimately be shown to be involved in this and a myriad other cases, it seems likely that at least part of such variation may be accounted for by a rule like this (dubbed ‘vowel harmony’ in Malone 1992b, but I now feel the regressive direction of the process more likely bespeaks umlaut):

- (15F) **Umlaut.** Under conditions remaining to be determined, the earlier V in VC(C)V assimilates in color to the later V.

Thus [lǎxta], [lǎxtow] ‘he took her, them’ ← *léxta*, *léxtow*; and while [léxtī] ‘he took him’ remains unchanged (C:232), note [gínβī] ‘he stole it’ (C:210) ← *génβī*. Note also unumlauted [hámbel], [hambélat] ‘he, she profaned’ vs. umlauted [hambīlīt] ‘I profaned’ (H:264–65), with the long [ī] possibly by 14F, and conversely by shortening 14F in [lǎ-déhlet] ‘let me not fear’ vs. unmarked [qǎ-dáhel] ‘he fears’ (C:211). (Rule 16F may in fact be iterative, if three-syllable stretches such as the following should be cases in point: [mahriβīlī] ‘they destroy it’ (C:219).)

Probably independently of 16F, there are also widespread [e ~ i] and [o ~ u] alternations (as in one form or the other common to most Northwest Semitic languages). Though one or two contributing factors are reasonably clear (such as the influence of *y* in favoring *i*, see forms under *√yhm* ‘sit’ at C:225), conditions are for the most part uncertain. However, some promising leads are adduced by Macuch at C:47.

10.3. Classical Mandaic

10.3.1. *Phonetic interpretation*

CM is written in a development of the so-called West Semitic syllabary notable as the only autochthonous Semitic orthography to have evolved from a (tendentially) consonant-only representational system to a fullblown functional alphabet bringing vowels up to parity with consonants without resorting to supernumerary diacritics. The strategy was wonderful for its simplicity. A handful of letters originally representing glides (notably *ʕ*, *w*, *y*, *ʔ*) were pressed into virtually obligatory double-duty service for presenting vowel (respectively *a*, *u/o*, *i/e*, *i/e*). In what follows these letters will be transcribed as <a, u, i, e>, answering to Macuch’s <a, u, i, ^c>. Though <i> and <e> both represent *i* or *e*, <i> is the all purpose letter for this, <e> being for the most part restricted to special positions, notably anlaut. Thus <ehbit> [ehβeθ] ‘I gave’ while initial <i> in <iahbit> represents nonvocalic *y*, [yahβeθ], also ‘I gave’ (D:189).

Ironically CM orthography, by the very virtue of becoming fully alphabetic, poses certain ambiguities of phonetic interpretation to a greater degree than some at best partially alphabetic sisters and cousins like Babylonian Talmudic Aramaic (BTA) and Mishnaic Hebrew (MH). This is due to

the fact that CM writing systematically fails to indicate segment length, V or C, while BTA and MH tend to use their glide letters (so-called ‘matres lectionis’) only (or largely) for long vowels: thus CM <basim> ‘fragrant’ (D:48) might represent adjectival [bassīm] or stative-participial [bāsem] and in either event would be homographic with factive [bassem] ‘make fragrant’ (cf. D:67).

Another pervasive ambiguity of CM orthography derives from the system’s failure to indicate spirantization (§ 10.3.3 below), in that <p> may stand for [p] or [f] (let alone [pp] and maybe even [ff]), and likewise <t; k; b; d; g> vis-à-vis [t, θ; k, x; b, β; d, ð; g, ɣ].

The two widest-ranging tools for dealing with these and other cataracts of CM orthography are the panoply of comparative Semitics and the testimony of MM. A third resource, despite its apparent richness, tends to be of rather limited help. I refer to the traditional reading pronunciation of CM texts by MM-speaking priests (*tarmīdī*), concerning reliance on which Macuch rightly cautions care (e.g. H:104). Based on my own experience, a rule of thumb has emerged favoring credence to MM over the traditional pronunciation whenever the two conflict. Despite my reservations however, the *tarmīdī*’s readings are not always bootless to the job of reconstruction; see e.g. § 10.3.3.

Considerable latitude of interpretation is often possible in reconstructing CM phonetics. In what might be called the **progressive mode** (p-mode), we would hypothesize CM to be as similar to MM as the overall reconstructive enterprise will allow. In the **conservative mode** (c-mode), on the other hand, we would take the opposite tack, and hypothesize CM to be as old-fashioned as possible, again subject to compatibility with the overall comparative-reconstructive frame.

In what follows, both modalities will receive some discussion, but only the conservative mode will be consistently developed.

10.3.2. Consonant groups

If we apply the bipolar reconstructive strategy mentioned in § 10.3.0 to initial C groups in CM, we may posit at the progressive pole that these distributed quite as in MM (§ 10.2.1), evincing all of [#CC-] ~ [#VCC-] ~ [#CVC-]; or alternatively at the conservative pole posit only [#ṼCC-] and [#CṼC-] for full-blown Cs, admitting #CṼC- only for initial ghost consonants (§ 10.4), as #GṼC- = [#ṼC-].

For non-geminate medial and final groups, however, CM contains resolutions (virtually) absent in MM: [-CṼCC-] and [-CṼC#]. These aside, we would have p-modal [-CCṼC-] (improbably also [-CCC-]), [-CCṼC-] (~ [-CCṼC-]), and [-CC#], as in MM; or c-modal [-CCṼC-] and [-CC#].

The ambivalence of proclitics seen for MM (independent word? word part?) probably holds for CM as well: thus the relative autonomy of the preposition in [men lelyā] ‘than (from) night’ (D:274) vs. its fusion in [mellebbay] ‘from my heart’ (D:237).

10.3.3. Rules for consonant groups

The alternations of the type considered in § 10.3.1 are accountable, in the p-mode, by antecedents of the three corresponding MM rules considered in § 10.2.2 (Reduction, Promotion, Prothesis) plus a rule of **Epenthesis**. In the c-mode, I will present more restricted versions of Reduction and Promotion. Let us start here with the former (the significance of the star in 17B* will be explained shortly):

(16B*)Reduction'.

- a. A non-long open-syllabic vowel is deleted when its syllable immediately follows a vowel, and weakened to *ə* elsewhere.
- b. Instrumentation. Regressive iterative (see also 20 below).
- c. Restriction. Does not apply to the first full vowel (i) in verbs of three or more syllables or (ii) in lexically specified nominals (Malone 1969, 1992a).

Thus *helaqā* → [helqā] ‘fate’ (emphatic state) (D:145), whereby regressivity dictates deletion of *a* prior to *e*, and iteration dictates that the rule scan its own output (which in this case fails to implement Reduction' of *e* since deletion has closed its syllable); *helaq* → [həlaq] ‘fate’ (absolute state) (D:145), contrast [hamar] ≠ [həmar] by restriction c.

Reduction' must follow Apocope':

(17A) Apocope'. With certain lexical exceptions, word-final vowels are dropped.

Thus *pərašū* → [pəraš] ‘they understood’ (H:348). A lexical exception is nominal *-ā*, as just seen in [helqā] ‘fate’.

The CM-specific alternants [-CṼCC-] (to [-CCṼC-]) and [-CṼC#] (to [-CC#]) (§ 10.3.1) arise through a rule of epenthesis (again, explanation of the star in a moment):

(18B*)**Epenthesis'**. With lexical exceptions, a group $-C(\check{V})C(\check{V})C-$ or $-CC\#$ is resolved as $-C\check{V}CC-$ or $-C\check{V}C\#$, where $\check{V}$ is unmarkedly *e* or (lexically determined) *e* ~ *a* respectively. (For the $\check{V}$ in $-C(\check{V})C(\check{V})C-$, see § 10.4.)

Thus *pərašūy* 17B* → *pəršūy* → [peršūy] 'they understood him' (D:381), *nešmoṭan* 17B* → *nešmōṭan* 19B* → [nešemṭan] 'may he draw me forth' (D:470), *nəšabatnnon* 19B* → *nəšabatennon* (→ *nəsaβaθennon* by 25B*, see § 10.3.3) → [nəsaβθennon] 'she took them' (D:302), *malk* 19B* → *malek* (25B* → *malex*) 17B* → [mālex] 'king' (D:244). (The most notable lexical exception to Epenthesis' keys on the subject suffix *-t* 'you', which bonds in auslaut to a preceding consonant: [axalt] 'you ate' (D:16).)

It will be noted that in the first two examples the order of application is Reduction' ^ Epenthesis' while in the last two it is Epenthesis' ^ Reduction'. This is due to the action of the so-called **Syllable Adjustment Circuit** (for more on this, in part tacitly revised here, see Malone 1992a):

(18) **The Syllable Adjustment Circuit**. Rules tagged with *, so far Reduction' (17B*) and Epenthesis' (19B*), interact to apply regressive-iteratively until the structural descriptions of all have been exhaustively satisfied.

CM schwa coloring has (approximately) the same content as its MM reflex:

(19C) **Schwa Coloring'**. $\check{V}$ may labialize (*ō*) or lower (*ā*) under the influence of an adjacent C-position.

Thus *šəmə* → [šōmə] 'he heard' (D:469).

Rule 21C in turn feeds CM Promotion', a process (by hypothesis) considerably more restricted than its MM descendant:

(20D) **Promotion'**. Initial $\check{V}$ promotes to $\check{V}$.

Thus *Gəxalt* 21C → *Gǎxalt* (for G as trigger, cf. 2D) 22D → [axalt] 'you ate' (D:16).

Like MM 2D, CM 21C and 22D are fed by a rule of prothesis:

(21B*)**Prothesis'**. An initial group $\#C(\check{V})C-$ may resyllabify to $\#\check{V}CC-$.

Also similarly to MM, Prothesis' is quite restricted in CM (though there is some evidence it may have been more vigorous in the interregnum Post-classical Mandaic). Example: [sǒfīhāōā] ~ [esfīhāōā] 'lips' (H:124).

Finally, CM Simplification' works out quite like its MM reflex (5B); it is ordered to follow the circuit (20):

(22C) **Simplification'**. (Same content as 5B)

Thus *šaddaran* 17B* → *šaddāran* 24C → [šadran] 'he sent me' (H:483). (Note that Epenthesis' (19B*) cannot preempt Simplification' to give impossible *šadedran*; see § 10.4.)

10.3.4. *Spirantization*

At the progressive pole, the phenomenon of spirantization in CM might approach that suggested for MM in § 10.2.3, to the extent that that situation was not catalyzed by the great influx of Persian and Arabic loans characteristic of MM.

At the conservative pole, we should have the well-known classical Northwest Semitic rule in all aspects short of its ordering with respect to the rest of CM phonology:

(23B*)**Spirantization'**. With possible lexical exceptions (notably /-t/ 'you'), a non-uvularized simplex stop (*p, t, k, b, d, g*) is spirantized (*f, θ, x, β, ð, ɣ*) postvocally.

The star marks this rule as participating in the Syllable Adjustment Circuit (20), along with Reduction', Epenthesis', and Prothesis' (17B*, 19B*, 23B*). By dint of the circuit, we may expect to find Spirantization' concurrently ordered both earlier and later than other circuit rules within one and the same form: e.g. /nekdoβī/ 25B* → *nekdoβī* 17B* → *nekðāβī* 19B* → *nekedβī* 25B* → *nekeðβī* 25B* → [nexedβī] 'he writes it' (D:204).

The evidence justifying the ordering of Spirantization' into the circuit is wispy indeed. Since the orthography (with rare and irrelevant exceptions) gives no clue as to when a spirantizable stop is in fact spirantized, the major vehicles of potential evidence must be MM and the traditional pronunciation of CM. As may likely be gathered from § 10.2.3, analogical change has all but effaced the original patterns of spirantization in Mandaic as a phonological rule. The analogical fundamentals of 8 are especially apparent, since in CM most tokens of 2√C and 3√C occurred postvocally, particularly in the simple (qal, peal) verb, while most of 1√C failed to. This in fact provides us with a tool for teasing out a reconstruction of the ordering conditions for 25 in CM, on the basis of the precious few traditional pronunciations relevant in this regard which defy 8. Thus the spirantal 1√f, 1√ɣ in Prothesized'

(23B*) t[eftahnīn] ‘we opened’, t[eȳtal] ‘he killed’ (reading <eptahnin> H:52 and <egtal> C:46) are plausibly interpretable to the effect that 25 could follow 23B* (pace t[ēptet] ‘you opened’ which obeys 8, for <eptit> H:30). Since moreover 17B* must be allowed to precede 23B*, and the entire network of comparative Semitics makes it virtually foregone that 17B* must at least in some forms have followed 25 (space precludes demonstrating this here), the resulting ordering 25[~]17B*[~]23B*[~]25 invests 25 as a member of the circuit; hence the tag B*.

On the other hand the anti-8 occlusion of 2√p in t[šoprā] ‘bird’ (H:183, C:39), from *šopparā*, suggests that 25B*, and by hypothesis the whole circuit, preceded Simplification' 24, thus tagged with C: *šopparā* 17B* → *šop-pārā* 24C → [šoprā].

10.3.5. Stress

While the CM stress placement rule on the p-mode hypothesis may have been like that of MM (12E), in the c-mode it would rather have been like that traditionally reconstructed for the closely related early Classical Syriac:

(24C) **Stress Placement'**. V → [+stress]/___C₀#; i.e., stress the last vowel in the word.

Thus from § 10.3.2: /helaqā/ 17B* → *helqā* 26C → [helqā] (C₀ = ∅); /helaq/ 17B* → *həlaq* 26C → [həláq] (C₀ = q); /Gəxalt/ 25B* → *Gəxalt* 21C, 26C → *Gəxált* 22D → [axált] (C₀ = lt).

10.3.6. Other alternations

Though in the p-mode alternations of quantity such as those discussed for MM in § 10.2.5 may have prevailed as early as CM, in the absence of firm evidence the c-mode position adopted in this chapter is that they did not.

Turning to alternations of vowel quality, CM indeed shows evidence for a precursor of MM 15A:

(25D) **Assimilatory Lowering'**. $\dot{V} \rightarrow [+low] / \text{___} 3v[+low]$. (Discursive statement as per 15A)

Thus from the /CaCCeC/ verb: *tabbér* ‘[tabbár] ‘he broke’ (D:482), *bašqér* → [bašqár] ‘he recognized’ (D:71) (contrast MM [bašqer] (§ 10.2.6)), and probably such as [šabbā] ‘he praised’ (D:446) from *šabbéh* followed by a rule changing $\dot{V}h^f$ to $\tilde{V}$ (cf. Malone 1985). ‘Exceptions’ to 27C such as [šabbəhōí] ‘I praised him’ (D:447) and [bašqerōí] (~ [bašqarōí]) ‘I recognized

him' (D:71) suggest that postfocal stress may render the rule variable, thus the tag D (Stress Placement' 26 being tagged C).

There is little if any evidence for umlaut (cf. 16F) in CM, though some spelling variations adduced in H:15 may be suggestive: e.g. <birikta> ~ <bu-rukta> 'blessing'.

Finally, the status of [e ~ i] and [o ~ u] in CM remains unclear. Interpretation at the c-pole leads me to posit consistent *e* and *o* (largely following Syriac); thus <diqna> as only [deqnā] 'beard' (D:109) despite MM [deqnā] ~ [diqna] (C:213, VV).

10.4. Appendix

For greater accessibility, the phonology fragments covered in this chapter have been largely presented in conservative conventional format. In recent years, however, great strides have been made in phonological theory, with no mean showing in Semitic. And while much of this latter work is rather directly in the tradition of John McCarthy's bellwether dissertation (1979)—e.g. Hoberman 1988—some also reflects other theoretical innovations; e.g. Lowenstamm 1991, Malone 1988, 1989a. In this appendix just a sampling of relevant Mandaic points from the body of the chapter will be touched on.

What are called above **ghost consonants** (§§ 10.2.2, 10.3.1) have been explicated as featurally empty melodic segments in Malone 1991. (The term 'ghost' is from Szpyra 1992, whither also the reader is referred for a theoretical update and review of the general literature.)

While the necessity of parenthesizing $\check{V}$ in the conventional segmental format of rules like 4C, 5B, 19B*, and 23B* ill reflects its hypothesized **inertness**, metrical phonology provides a plausibly natural account (for Tiberian Hebrew, see Churchyard 1992).

Lastly, the failure of Epenthesis' (19B*) to change forms like *šaddšran* to the like of *šadedran* can be made to follow from the so-called **integrity** of geminates (see Goldsmith 1990 plus references, but note the caution sounded in Malone 1989b).

References

- Chomsky, N. & M. Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

- Churchyard, Henry. 1992. "The Tiberian Hebrew Rhythm Rule in the Typology of Rhythm Rules." Meeting of the Linguistic Society of America, Philadelphia, January.
- Drower, E.S. & R. Macuch. 1963. *A Mandaic Dictionary*. London: Oxford University Press.
- Goldsmith, John A. 1990. *Autosegmental and Metrical Phonology*. Oxford: Blackwell.
- Hoberman, Robert. 1988. "Local and Long-distance Spreading in Semitic Morphology." *Natural Language and Linguistic Theory* 6.541–49.
- Lowenstamm, Jean. 1991. "Vocalic Length and Centralization in Two Branches of Semitic." *Semitic Studies in Honor of Wolf Leslau on the Occasion of his Eighty-Fifth Birthday*, ed. by Alan S. Kaye. Wiesbaden: Harrassowitz.
- Macuch, Rudolf. 1965. *Handbook of Classical and Modern Mandaic*. Berlin: de Gruyter.
- . 1989. *Neumandäische Chrestomathie mit grammatischer Skizze, Kommentierter Übersetzung und Glossar*. Wiesbaden: Harrassowitz.
- Malone, Joseph L. 1969. "Rules of Synchronic Analogy: A Proposal Based on Evidence from Three Semitic Languages." *Foundations of Language* 5.534–59.
- . 1976. "Messrs Sampson, Chomsky and Halle, and Hebrew Phonology." *Foundations of Language* 14.251–56.
- . 1985. "Classical Mandaic Radical Metathesis, Radical Assimilation and the Devil's Advocate." *General Linguistics* 25.92–122.
- . 1988. "Lexical Phonology and the Aramaic Verb of the Onkelos and Jonathan Targums." Meeting of the American Oriental Society, Chicago, March.
- . 1989a. "Prosodic Domains for Tiberian Hebrew Phonology." North American Conference on Afroasiatic Linguistics, New Orleans, March.
- . 1989b. "Geminates, the Obligatory Contour Principle, and Tier Conflation: The Case of Tiberian Hebrew." *General Linguistics* 29.112–30.
- . 1991. "Underspecification and Phonological Assignment of Phonetic Strings: The Case of Classical Mandaic [qen:a:] 'nest'." *Actes du XIIème Congrès International des Sciences Phonétiques*, vol. 2. Aix-en-Provence: Université de Provence. Pp.130–33.

- . 1992a. “Diachronic–Synchronic Dystony: A Case from Classical Mandaic.” *General Linguistics* 32.37–58.
- . 1992b. Review of Macuch 1989. *JAOS* 112: 339–40.
- . 1993. *Tiberian Hebrew Phonology*. Winona Lake: Eisenbrauns.
- McCarthy, John J. 1979. “Formal Problems in Semitic Phonology and Morphology.” M.I.T. Ph.D. dissertation.
- Szpyra, Jolanta. 1992. “Ghost Segments in Nonlinear Phonology: Polish yers.” *Language* 68.277–312.

This page intentionally left blank.

Chapter 11
Old South Arabian Phonology

Gene Gragg
University of Chicago

11.1. Introduction

Other chapters in this volume present detailed material on a set of Semitic languages and language subfamilies, Arabic, Modern South Arabian, and Ethiopian Semitic, which geographically occupy the southern part of the Semitic language distribution, and which are often grouped together under the rubric “South Semitic.” In addition to these languages, in the core of this southern Semitic area, in the modern Republic of Yemen, from early in the first millennium B.C.E. to roughly the rise of Islam, a large number of monumental inscriptions attest to the existence of Old South Arabian (OSA), a group of closely related languages distinct from any of the other representatives of “South Semitic.” There is considerable debate in the Comparative Semitic literature whether all these languages in fact constitute a sub-family of Semitic in the strict sense: a group of languages which derive from a distinct ancestral speech community, and which share a period of common development not participated in by other attested Semitic languages. The opposing hypothesis is that some or all of the similarities among these languages are the result of parallel development from Common Semitic, with tendencies toward convergence reinforced by (more or less direct) chains of (more or less continuous) contact among speech communities which have been geographically contiguous for a millennium or more.

This chapter summarizes the basic facts of OSA phonology (authoratively presented in Beeston 1962, 1984) and situates OSA with respect to the other “South Semitic” languages.

11.2. Old South Arabian

The number of published Old South Arabian (also referred to as Epigraphic South Arabian) texts runs into the thousands, and more are being published each year. See CIASA (1977) and Avanzini (1977, vol. 1) for bibliography up to 1974; more recently note Beeston et al. (1982) and Robin (1992). The

vast majority of the inscriptions are in one of four languages, corresponding to the four city-states/kingdoms located on a line running from northwest to southeast, from the Saudi Arabian border to eastern Yemen, between the inland slope of the Yemeni highlands and the Arabian desert—that is, along the southern end of the “spice route” between the Indian Ocean and the Mediterranean. These are, from north to south, the languages of: Maʿin, Sabaʾ, Qaṭabān (or, more properly, Qīṭban), and Ḥaḍramawt. Of these Sabaic, the language of Sabaʾ, is by far the best attested. After the first centuries B.C.E. the political center of Yemen shifted to the highlands (in connection with the political ascendancy of the Himyarites, and perhaps also with a shift of an important part of the spice trade to the Red Sea), and inscriptions were only drawn up in Sabaic, perhaps no longer a spoken language.

In spite of the relative abundance of documents, our knowledge of OSA is limited by a number of factors:

1. the highly stereotyped content of the monumental inscriptions;
2. a discourse style which consistently avoids using the first or second person—we know very little about the realization of these categories in the complex inflectional morphology (but see now Ryckmans et al. 1994);
3. a writing system with no indication of short vowels or consonantal gemination, no indication of /ā/, and only irregular indication of other possible long vowels, except, probably, word-final /ī/ and /ū/.

Nevertheless several things are certain. One is that the twenty-nine characters of the OSA alphabet, presented in Table 11-1 in their usual transliteration values, do seem to correspond to the twenty-nine reconstructable consonants of Proto Semitic.

11.2.1. *Comments*

11.2.1.1. *Emphatics*

On the basis of evidence from Modern South Arabian and Ethiopian Semitic (q.v.), it is likely that the “emphatics” were glottalized in OSA, and not pharyngealized, as in Arabic.

11.2.1.2. *Labial (f)*

A Latin transcription insures that the labial consonant in OSA was /f/ as in MSA and Ethiopian Semitic, not /p/ as in Canaanite, Aramaic, and Akkadian.

Table 11-1. Old South Arabian Consonants

	Labial	Intdnt	Dent	Palatal	Lar'al?	Velar	Phar	Glottal
Stop								
Vceless			t			k	ʕ	ʔ
Voiced	b		d			g		
Emph			ṭ			q		
Continuant								
Vceless	f	ṯ	s ³	s ¹	s ²	ħ	ḥ	h
Voiced		ḏ	z			ʁ		
Emph		ẓ	ṣ		ḍ			
			r		l			
	m		n					
	w			y				

11.2.1.3. Interdental (z)

Although this shows a regular correspondence with what is commonly taken to be a voiceless interdental emphatic spirant /θ/ in Proto-Semitic, its grapheme 𐩦 closely resembles the grapheme for /š/ 𐩣, and in many texts it seems to have merged with /š/ (note common modern Arabic pronunciations).

11.2.1.4. Velars (k, g, q, ħ, ʁ)

/q/ could have been simply a voiceless glottalized velar [kʰ]—as in Ethiopic Semitic and MSA. Note however the uvular articulation in Arabic. Note also that /ħ, ʁ/ are post-velar spirants in both Arabic and MSA (they have not been preserved in modern Ethiopic Semitic or in the pronunciation tradition of Geʿez).

11.2.1.5. Sibilants and laterals (s¹, s², s³, ḏ)

The transcription with “s” and superscript numbers has been widely adopted for the graphemes 𐩣 = s¹, 𐩥 = s², 𐩦 = s³, largely because of the generally inconsistent and ambiguous use of the symbols “s”, “š”, and “s” in the older literature. /s³/ corresponds to Proto-Semitic /*s/ in etymologies and in loanwords and /s¹/ to Proto-Semitic /*š/. /s²/ corresponds to a Proto-Semitic consonant which is preserved as a separate phoneme only in MSA, where it is realized as a voiceless lateral spirant. Elsewhere in Semitic it merges with

either /š/ or /s/, although it is preserved graphically in the Hebrew “*śin*” (see Steiner, 1977, as well as Tables 11-2 and 19-1). Historical and comparative evidence suggests that /d/ was the voiced and/or “emphatic” counterpart to /s²/. Note that the occasional use of “*s¹*” for “*s³*” in the late OSA texts may indicate a merger of these two consonants at this period, as happened in Arabic. In Ḥaḍramawt /s³/ and /t/ seem to have merged, with their graphemes used interchangeably.

11.2.1.6. *Glides (w, y)*

These occur in all positions. The graphemes “*w*”, “*y*” in final position apparently can represent the, possibly long, vowels /u, ū, i, ī/, as in *s¹trw* /šaṭāru/ ‘they wrote’, *-hw* /-hu/ ‘his’, *-hmw* /-hVmu/ ‘their (masc.)’—note the possible writings *-h*, *-hm* for the latter. In medial position they can probably also represent vowels, as well as diphthongs or independent consonants: thus *kwn* ‘he was’ (which has *kn* as an alternate shape) is presumably /kona/, although /kawna/ cannot be absolutely excluded (for a different lexical item /ka(w)wana/ could also have been a possibility).

11.2.2. *Phonological processes*

Beeston (1984: 10f.) lists a number of phonological processes of the expected sort in a Semitic language: loss of laryngeals /ʔ, ʕ, h/ in certain positions; fluctuation of /w, y/; assimilation of /n/ to following consonant (not attested before /ʔ, ʕ, h, ḥ, ʁ, m/); assimilation of /d/ to following /t, t/ in numerals ‘one (fem.)’, ‘six’, ‘sixty’ *ʔḥdt > ʔḥt, s¹dθ > s¹t, s¹dθy > s¹ty. However, absence of real information about vocalic environment make a precise evaluation impossible. A replacement of earlier s²ltt ‘three’ by *ltt* in later Sabaic may be an instance of assimilation at a distance.

11.2.2.1. *Phonological variants in OSA grammatical forms*

A well-known major morphological isogloss within Semitic involves the use of /š/ or /h, h > ʔ/ for the causative formative with verbs and the base of the third person pronoun (Akkadian: *šuprus-* ‘cause to decide’, *-šu* ‘his’; Ugaritic: *šlhm* ‘cause to eat, feed’, *-h* ‘his’; Hebrew: *hiktīb* ‘cause to write’, *hu* ‘his’). Along the same lines, chapter 19 points out a number of /š ~ h/ alternations involving these and other morphological and lexical contexts, within and between various MSA languages. Whatever is the ultimate phonological or morphological explanation of this alternation, it is interesting to note that **within** OSA, the Sabaic dialect shows the /h/ alternate for both the causative

and pronominal morphemes (*hqny* ‘dedicate; cause to possess’, *-h* ‘his’), while Ma‘in, Qīṭban, and Ḥaḍramawt have *sʰ* in both cases (*sʰqny*, *-sʰ*).

A number of inner OSA dialect variations involve a “parasitic,” non-etymological /h/, which appears in a variety of contexts in the non-Sabaic languages: in certain lexical forms in Ma‘in (*bhn(t)* ‘sons (daughters)’, *ṭhmn* ‘eight(y)’, *lhm* ‘not’); as a noun suffix in construct singular and broken plural genitives, and in construct duals and external plurals in Ma‘in (*bn ṛydw-h-sʰm* ‘from hands-h-their’) and occasionally in Qīṭban and Ḥaḍramawt; in a possible “adverbial accusative” *-h(m)* in Ma‘in (*sʰlmhm* ‘in peace’) and Ḥaḍramawt; in a determinate singular and broken plural suffix *-hn* in Ḥaḍramawt, corresponding to *-n* in Sabaic and elsewhere. While at least some of these instances may involve phonological rather than morphological innovation not enough is known about the suprasegmental and vocalic environment to allow the formation of any coherent hypotheses.

In Qīṭban there is a regular alternation between a short form and a long form of the third person singular suffix pronoun: masculine *-sʰ* ~ *-sʰww*; feminine *-sʰ* ~ *-sʰyw*. The short form occurs with singular and internal plural forms, the long form with dual and external plural forms. Both pre-suffix vowel length and pre-suffix stress have been suggested as a possible conditioning factor (Beeston 1962: 45 observes that this might correlate with certain of the parasitic /h/’s just noted)—but again, hard evidence is lacking. Ḥaḍramawt masculine singular pronouns seem to show the same alternation. In Ma‘in an occasional masculine singular variant *-sʰw*, in place of the much more frequent *-sʰ*, might also be an indication of a similar phenomenon, although these are not enough attestations to be certain.

11.3. South Semitic

In the first place it should be noted that many of the criteria used to define South Semitic are more morphological than phonological in nature. The most important are: [1] The existence of a highly developed system of internal plurals (of the kind Arabic: *bayt* ~ *buyūt*, Ge‘ez: *bet* ~ *abyāt*, Soqotri *beyt* ~ *ībihet* ‘house ~ houses’). [2] Presence of a bi-syllabic present stem of the form CV(C)CVC, as in Ge‘ez *yə-qattəl* and Soqotri *yə-k’abər* (versus CCVC as in Hebrew *yi-qtol*). [3] Forms with /k/ in the first and second person subject suffixes of the past tense as in Ge‘ez *-ku*, *-ka*, *-ki* (for 1sing., 2masc.sing., 2fem.sing.) and Soqotri *-k*, *-k*, *-š* (< *-ki*) (versus forms with /t/ as in Hebrew *-ti*, *-tā*, *-t*). With respect to criterion [1], OSA clearly belongs with the rest of

“South Semitic” (Sabaic *byt* ~ *?byt*). But for [2] and [3] Arabic (which shows *ya-qtul-u* and *-tu, -ta, -ti*) clearly patterns with Hebrew and Aramaic. On the other hand, for OSA, while [2] is still unknown because of lack of information about short vowels and gemination, for [3] emerging evidence (clear second person *-k* forms in Ryckmans et al. 1994; note also a possible form *rhmk* ‘thou art merciful’ cited in Beeston 1984: 14) seems to indicate that OSA might pattern with MSA and Ethiopic.

Table 11-2. Semitic Consonant Correspondences

PSem	Akk	Ugar	Aram	Heb	Arab	Ge‘ez	OSA
*p	p	p	p	p	f	f	f
*θ	š	ṯ	t	š	ṯ	s	ṯ
*ð	z	ḏ	d	z	ḏ	z	ḏ
*ð̣	ṣ	ṯ̣	ṯ̣	ṣ	ṯ̣	ṣ	ṣ̣
*š	š	š	š	š	s	s	s ¹
*ṯ	š	š	s	“š” > s ^a	š	“š” > s ^a	s ²
*ḏ	ṣ	ṣ	ṯ̣	ṣ	ḏ	“ḏ” > ṣ ^a	ḏ
*ḥ	ḥ	ḥ	ḥ	ḥ	ḥ	“ḥ” > ḥ ^a	ḥ
*ʕ	∅	ʕ	ʕ	“ʕ” ^b	ʕ	ʕ	ʕ

a. Earlier graphemic distinction not reflected in pronunciation tradition.

b. Evidence from transcriptions for ʕ ≠ ʕ̣.

The phonological evidence is summed up in Table 11-2, which gives the most common non-identity consonant correspondences for Proto-Semitic, Akkadian, Ugaritic, Aramaic, Hebrew, Arabic, Ge‘ez, and OSA (minus the cases where Akkadian has simply dropped a laryngeal—see chapter 19 for the sibilant correspondences in MSA). It is immediately obvious that, as already mentioned, OSA preserves intact all of the consonantal segments of Proto-Semitic, as does MSA. Arabic loses only one consonantal distinction, through a process whereby /š/ merged with /s/, after which /ʔ/, the voiceless lateral spirant, became /š/. However, apart from the shared /f/ in Arabic, OSA, and Ge‘ez (also in MSA), as opposed to /p/ elsewhere (frequently cited as a feature of South Semitic), there is little other evidence of the kind of shared phonological innovation that would imply a common sub-family node and a hierarchical family tree structure. One case one might point to,

the common treatment of Proto-Semitic /ħ/ and /ʕ/ in Arabic and Ge'ez, fails to lead to an unambiguous conclusion, since it is contradicted by the morphological evidence cited above.

It is in fact hard to escape the conclusion that the major phonological developments in Semitic, converging or not, took place independently in each of the major branches, and that each branch seems to have started its independent existence with the Proto-Semitic consonant system virtually intact. In this context one might note that even the "South Semitic /f/" might reflect a more complicated Pre- and Proto-Semitic situation than the postulation of a simple spirantization of Proto-Semitic /p/ might suggest (cf. widespread Afroasiatic /f/ for Semitic /p = f/: Semitic **pu*- 'mouth' = Cushitic **af*; Semitic **ḫVpr*- 'nail, claw' = Berber *a-tfər*).

The upshot of all this seems to be that Ethiopian Semitic and MSA indeed form a sub-family of Semitic. It would be geographically and historically plausible that OSA would also belong to this family, and there is a small amount of linguistic evidence pointing in this direction. But this evidence is not detailed enough to tell us what the relation of OSA is to the other two, and what the resulting family substructure might be. Arabic belongs to a linguistic and cultural South Semitic contact area, whose nature still needs to be worked out. While Arabic certainly interacted with the other South Semitic language groups, in certain crucial respects Arabic seems to have followed paths of development not participated in by "South Semitic" proper.

Bibliography

- Avanzini, Alessandra. 1977–80. *Glossaire des inscriptions de l'Arabie du sud 1950–1973* (Quaderni di Semitistica 3), vol. 1 *Bibliographie*, vol. 2 *Glossaire*. Florence: Istituto di Linguistica e di Lingue Orientali.
- Beeston, A. F. L. 1962. *A Descriptive Grammar of Epigraphic South Arabian*. London: Luzac.
- . 1984. *Sabaic Grammar* (Journal of Semitic Studies Monograph 6). Manchester: University of Manchester.
- Beeston, A. F. L., M. A. Ghul, W. W. Müller, and J. Ryckmans. 1982. *Dictionnaire Sabéen (anglais-français-arabe)*. Louvain-la-Neuve: Peeters.
- Biella, Joan. 1982. *Dictionary of Old South Arabic: Sabaean Dialect* (Harvard Semitic Studies 25). Chico, Calif: Scholars Press.

- CIASA. 1977–. *Corpus des inscriptions et antiquités sud-arabes*. Louvain-la-Neuve: Peeters. [Gives authoritative editions, copies, photographs, translations (usually in French or English) of more recently discovered texts. Continues several earlier series of text publications, in particular the *Corpus Inscriptionum Semiticarum* [CIS, CIH] part 4 (Paris, 1889–1929) and the *Répertoire d'Epigraphie Sémitique* [RES] vols. 5–8 (Paris, 1929–1968).]
- Jamme, Albert. 1962. *Sabaeen Inscriptions from Maḥram Bilqīs (Mârib)*. Baltimore: Johns Hopkins University Press.
- Ricks, Stephen. 1989. *Lexicon of Inscriptional Qatabanian* (Studia Pohl 14). Rome: Editrice Pontificii Istituto Biblico.
- Robin, Christian. 1992. *Inventaire des inscriptions sudarabiques*, 2 vols. Paris: Boccard.
- Ryckmans, Jacques, W. W. Müller, and Y. M. Abdullah. 1994. *Textes du Yémen antique inscrits sur bois*. Louvain-La-Neuve, Belgium: Institut Orientaliste.
- Steiner, Richard C. 1977. *The Case for Fricative-Laterals in Proto-Semitic* (American Oriental Series 59). New Haven: American Oriental Society.

Chapter 12
Ge'ez Phonology
Gene Gragg
University of Chicago

12.1. Introduction

Ge'ez (*gəfəz*), a Northern Ethio-Semitic language, was the official language of the Axumite kingdom which occupied large parts of present-day highland Eritrea and the Northern Ethiopian province of Tigre during much of the first millennium A.D., reaching the height of its power and regional influence between the fourth and sixth centuries. Ge'ez is first attested in pre- and early-Axumite inscriptions from the first centuries A.D., and is seen in near-classic form in a corpus of monumental inscriptions associated with a ruler named 'Ezana, and usually dated to the fourth century. 'Ezana's inscriptions also attest to the appearance of Christianity on the Ethiopian highland; and subsequent centuries witness the appearance of a Bible translation into Ge'ez, and the nucleus of what gradually became a vast ecclesiastical, civic, and court literature, continually accumulating during the long period in which Ge'ez was the only official written language of the Ethiopian church and state. Up to the threshold of the twentieth century, some knowledge of Ge'ez was a prerequisite for both civil and ecclesiastical careers in Ethiopia; and Ge'ez is still today the liturgical language of the Ethiopian Orthodox Church. As a consequence, a characteristic feature of Ethiopian civilization has been an extensive monastery- and church-based educational tradition, formed around a curriculum leading to a mastery of Ge'ez.

Within Northern Ethio-Semitic, Ge'ez is closely related to modern Tigre (northern highlands and Red Sea coastal plain), and stands in a more or less proximate ancestral relationship to Tigrinya, which probably started to emerge as a distinct entity on the home territory of Ge'ez from around the tenth century on. The remaining Ethio-Semitic languages (a dozen or so, including Amharic) belong to a separate (Southern) group, which cannot be derived from any attested Northern Ethio-Semitic language. Northern and Southern Ethio-Semitic, however, do seem to constitute a distinct genetic node in the Semitic family tree, a node which is most closely related to a Southern group of Semitic languages that includes Epigraphic and Modern

South Arabian. The exact historical relationships among Epigraphic South Arabian, Modern South Arabian, and Ethio-Semitic, as well as the relation of these three families to (North) Arabic, have been difficult to establish because of what was until quite recently insufficient data on the Modern South Arabian languages, and because of the phonological indeterminacy and morphological poverty of the textual evidence for Epigraphic South Arabian (an extensive corpus, but written in one of the more resolutely vowelless of the Semitic writing systems, and in a discourse format which, in spite of a respectable diversity of subject matter, managed to restrict itself almost entirely to third person pronominal and verbal forms).

12.2. The evidence for Ge'ez phonology

Two lines of evidence have to be correlated in drawing any conclusions about the phonology of Ge'ez in its "classical" (i.e., pre-tenth century) form. One is the writing system itself, together with the inferences suggested by related Ethiopic and non-Ethio-Semitic linguistic evidence. The other comes from the pronunciation tradition developed in the context of the traditional educational system mentioned above.

As far as the Ethiopic writing system is concerned, its core, the representation of the consonantal inventory, is already present in the earliest pre- and early-Axumite inscriptions. This consonantal script is derived more or less directly from some version of the South Semitic writing system which also appears in the Epigraphic South Arabian inscriptions. The major missing element, the representation of the vowels, is provided by a major innovation introduced in the later inscriptions of 'Ezana. Alone among the Semitic scripts, Ethiopic represents vowels, not by a separate set of (super-linear or sub-linear) vowel signs, but by means of a more-or-less uniform modification of the basic letter-shape. The base form of the consonant sign (the so-called 'first order' form) is taken to represent the consonant followed by /ä/ [æ], and six other alterations of the basic shape are introduced to represent the consonant followed by the vowels /i u a e ə o/ (the second through seventh orders; note that the sixth order is used in addition to represent the consonant in isolation, with no following vowel). As an example, the result of this transformation for the consonant /b/ is:

ORDER:	1	2	3	4	5	6	7
SIGN:							
VALUE:	<i>bä</i>	<i>bī</i>	<i>bū</i>	<i>ba</i>	<i>be</i>	<i>bə, b</i>	<i>bo</i>

Each sign thus represents a CV sequence, and a twenty-six member alphabet is thereby transformed into a 182 member syllabary (to which are added five forms each of the consonants /k g q ɣ/ to represent the labiovelars /k^w g^w q^w ɣ^w/ plus the vowels /ä i a e ə/, for a grand total of 202 signs).

The resulting writing system provided an unambiguous representation of Ge'ez words (and continues to provide the same service for Amharic and Tigrinya) except for two aspects of phonological shape—apart from stress, frequently not noted in practical orthographies in any case. These two problematic aspects are: (1) lack of a way of indicating the phonologically prominent and morphologically important feature of gemination (a two-dot diacritic, introduced to represent gemination of consonants in European grammars and dictionaries of Ethiopic since the seventeenth century, never became part of the manuscript or printed Ethiopic orthographic tradition); (2) lack of an unambiguous way of representing a consonant *not* followed by a vowel, i.e. in word- or syllable-final position (note the parallel ambiguity inherent in the Hebrew shwa symbol). As a consequence an orthographic representation <k₁l₆b₆> (where C_n is the *n*th order shape of the consonant C) might conceivably stand for any of the following values:

<i>kälb</i>	<i>käləb</i>	<i>kälbə</i>	<i>käləbə</i>	<i>källb</i>	<i>källəb</i>
<i>källəbə</i>	<i>kälbb</i>	<i>käləbb</i>	<i>käləbbə</i>	<i>källəbb</i>	<i>källəbbə</i>

(excluding values with /kk llbb/ on general phonotactic grounds).

Thus, for these missing features our only source of information is the other line of evidence, the pronunciation tradition. This tradition in its present manifestation is apparently fairly uniform, and seems to represent prestige pronunciation of Ge'ez in the central Ethiopian plateau (thus in a largely Amharic milieu) where royal residences and many centers of ecclesiastical influence have tended to be located since the decline of Axum. This pronunciation tradition is thoroughly Amharicizing in its treatment of consonantal values (but note the claim of Ullendorff 1955 about the existence of more "Tigrinya-izing" pronunciation patterns in the north). Still, the traditional pronunciation preserves stress, gemination, and syllable structure patterns which are at least in part distinct from those found either in Amharic or in Tigrinya, and as such may well reflect the state of affairs in an earlier stage of Ge'ez itself. In any case European scholarship from the seventeenth century on has been explicitly dependent directly or indirectly on information provided by a succession of Ethiopian literati who are acknowledged, usually by name, as the source of information for these aspects

of lexical and grammatical representations. (On pronunciation, cf. most recently Makonnen Argaw 1984; basic earlier works are Ludolph 1661, Trumpp 1874, König 1877, Littmann 1918, Cohen 1921, and Mittwoch 1926; the fundamental reference grammar, Dillmann 1857, 1907, relies essentially on Ludolph, with corrections from Trumpp and König included in the second edition by Bezold; Lambdin 1978, the most accessible pedagogical grammar, incorporates material from the traditional pronunciation, but without much explicit discussion; note in this context the section on *sāwasəw*, indigenous grammatical texts, by the Ethiopian scholar Getatchew Haile in Leslau 1987: xv–xvii.)

In the sections that follow, the representation of Ge'ez will thus be of a composite nature. Consonants and vowels (apart from the sixth order) will be given in a conventional transcription of the orthography (see the relevant sections below for interpretation or probable interpretation); representation of gemination and syllable structure (as well as stress, when indicated) will be based on what is known from the pronunciation tradition. Note that although the basic features of the pronunciation tradition are relatively clear, its scholarly investigation and evaluation is still under way. Note also, as far as the orthography is concerned, that apart from the relatively small corpus of Axumite inscriptions, apparently from the early and formative period of classical Ge'ez, almost all Ge'ez texts were either produced in a period when Ge'ez was no longer a spoken language, or are preserved in a long, poorly studied manuscript tradition, with a gap of many centuries between the period of formation of the core classical corpus (Bible translation, key liturgical, hagiographic, and monastic texts—perhaps sixth century) and the oldest extant manuscripts (rarely older than the fourteenth century). Thus the composing and transmission of texts in Ge'ez orthography are themselves aspects of the same process of cultural transmission which yields the pronunciation tradition.

12.3. Phonological segments

12.3.1. *Consonants*

In Table 12-1, row 1 gives the usual transcription value for each Ge'ez consonant, row 2 its traditional pronunciation, row 3 the corresponding consonant in Tigrinya, row 4 the most usual corresponding consonant, if present, in Arabic, and row 5 the first order shape of the corresponding consonant sign.

Table 12-1. Ge'ez Consonant Correspondences

p	ḥ	b	t	ṭ	d	k	q	g	f	s	ś	ṣ	ḍ	z	ḥ	ḥ	h	ʕ	ʔ	m	n	r	l	w	y
p	p'	b	t	t'	d	k	k'	g	f	s	s	ṣ	ṣ	z	h	h	h	ʔ	ʔ	m	n	r	l	w	y
p	p'	b	t	t'	d	k	k'	g	f	s	s	ṣ	ṣ	z	ḥ	ḥ	h	ʕ	ʔ	m	n	r	l	w	y
—	—	b	t	ṭ	d	k	q	g	f	s,θ	ś	ṣ	ḍ,z	z,ḏ	ḥ	ḥ	h	ʕ,ʔ	ʔ	m	n	r	l	w	y
T	ḥ	ḥ	ṭ	ṭ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ	ḍ

Details of interpretation are given below, but it is important to note the following: The traditional pronunciation gives essentially their Amharic value to the consonantal signs—pharyngeals /ʕ/ and /h/ merge with /ʔ/ and /h/ respectively, <ṣ> and <ḍ> are both pronounced /ṣ/, and <s> and <ś> are both pronounced /s/. Tigrinya (as well as Tigre) preserves the distinction between /ʔ/ and /ʕ/ and between /ḥ/ and /h/. But *no Ethiopic language* and *no element of the pronunciation tradition* provides the least bit of information about the pronunciation of the graphemes /ś/, /ḍ/, and /h/. Nevertheless, since the earliest pre-Axumite writing system did adopt these consonant signs from the parent South Semitic alphabet, while excluding a number of other signs representing consonants which had already merged in Ethiopic, they must have represented distinct consonants in early and classical Ge'ez. It is not clear when the mergers took place. The graphemes seem to be used consistently in the 'Ezana inscriptions, and in some strands of the manuscript tradition (recall, however, what was already said concerning the primitive state of the study of this tradition). However, variant writings of the same word with <ḥ> and <ḥ> begin to appear already in some late monumental inscriptions, and in the low end of the manuscript tradition, <ḥ/ḥ> (and sometimes <h>), <ḍ/ṣ>, and <ś/s> are used as virtual allographs. In Table 12-2, the consonants corresponding to <ś>, <ḍ>, and <ḥ> are interpreted with the help of data from cognate languages (note Arabic correspondences in Table 12-1).

With the foregoing reservations, the articulatory features of Ge'ez consonants will be interpreted as in Table 12-2.

12.3.1.1. *Emphatics* /ḥ, ṭ, q, q', s, ḍ/

As can be seen from Table 12-2, the phonological feature commonly termed "emphasis" in Semitic is implemented as glottalization in the pronunciation tradition (hence, Amharic) and Tigrinya. In fact the feature has this implementation in all Ethio-Semitic languages—and hence presumably also in Ge'ez. Since the Arabic implementation of this feature, velarization or

Table 12-2. Ge'ez Consonant Articulation

	Labial	Dental	ʔ	ʔ	Velar	Labio- velar	Pharyn- geal	Glottal
Stop								
Voiceless	p	t			k	k ^w	ʔ	ʔ
Glottalized	p̥	t̥			q	q ^w		
Voiced	b	d			g	g ^w		
Continuant								
Voiceless	f	s		ś	ħ	ħ ^w	ħ	h
Glottalized		s̥	q̥					
Voiced	z							
Nasal	m	n						
Sonorant		r l						
Glide	w	y						

pharyngealization, was once thought to be representative of early Semitic, glottalization in Ethio-Semitic was thought to represent the influence of language contact with earlier Cushitic languages in Ethiopia. However, the relatively recent discovery that emphatic consonants are also glottalized in the Modern South Arabian languages makes it possible that this might be a common South Semitic feature, and perhaps even common Semitic. Note in particular that /q/, which can be a uvular stop in some varieties of Arabic, is clearly a glottalized velar in Ethiopic. Note also that, because of the well-known difficulty in building up the glottal pressure necessary for a glottalized release while articulating a continuant like /s/, /ʃ/, the glottalized version of /s/, both in the pronunciation tradition and in modern Ethio-Semitic languages, tends to be realized phonetically as an affricate [ts]. It is uncertain whether the same was true of /q̥/. However, since it merged with the continuant /ʃ/, it is likely that it had become a continuant at least by the time of the merger.

12.3.1.2. Voiceless labials /p, ʔ, f/

Ge'ez is unique among the Semitic languages in having not only a stop and continuant voiceless labial consonant, but an emphatic labial stop as well. Akkadian, Aramaic, and Canaanite have only a /p/, while Arabic and other

South Semitic have only a /f/. Ge'ez /f/ is clearly what is cognate with the voiceless labial consonant of other Semitic, and Ge'ez thus patterns historically, as expected, with South Semitic and Arabic. A large number of the occurrences of /p/ and /p̥/ are in loanwords, mostly from or by way of Greek: *pilās* 'temple, gate' < Greek *pylē* 'gate', *pāppās* 'metropolitan, patriarch' < Greek *papās* 'father, title of priests', *pāpera*, *pāpera* 'purple' < Greek *porphyra* 'purple dye' (note that it is not always predictable when the consonant corresponding to Greek /p/ will be glottalized /p̥/ and when simple /p/—an indeterminacy common with other voiceless stops in loanwords from Greek). However, there are a large number of other occurrences, fully integrated into the native grammar and vocabulary of Ge'ez, where the origin of the stop is much less clear: *hepā* 'strike, throw, shoot with an arrow' (a so-called middle-weak verb from a root *hyṗ*, with a glide as a middle radical—on possible connection with Arabic *habba* consult Leslau 1987); *hāppālā* 'wash clothes' (Leslau 1987 notes Arabic *wabīl* 'stick with which one beats clothes that are being washed'), *qālāpā*, *qālṗāpā*, *qārṗāpā* 'chew, bite' (meanings pattern with verb *qālābā*, which may be borrowed from Amharic *qällābā*).

12.3.1.3. Labiovelars /kʷ, qʷ, gʷ, ḥʷ/

Ethio-Semitic is likewise unique among the Semitic languages for having developed a distinctive labiovelar series of consonants. They are already attested in the voweled 'Ezana texts (since they are represented by modifications of the basic consonant, there is no way a non-voweled text could indicate their presence). Labiovelars can occur before the non-rounded vowels /i e ä a ə/; their non-occurrence before /u o/ is undoubtedly a neutralization effect, since all consonants in the traditional pronunciation of Ge'ez (and in most modern Ethio-Semitic) have labialized allomorphs before /u, o/. Labiovelars are relatively widespread in the Ge'ez lexicon. A count of the letter G in Leslau 1987 reveals 259 entries with initial /g/, and 92 with initial /gʷ/. Where their origin can be traced, labiovelars seem to result from phonologization of labialized allomorph of a velar before a rounded vowel. Thus the verb *tārgʷāmā* 'interpret, translate' is reformed from an Aramaic word *targūm* 'interpretation' (note that this form of the word is not attested in Ge'ez, where the various words for 'interpretation', *tərgʷame*, *tərgʷəmt*, *tərgʷəmənnā*, are all nominalized forms of the Ge'ez verb). Likewise the labiovelar in the Ge'ez word for 'brother' *əḥʷ* is related to the fact that in many Semitic languages the word 'brother' has forms showing a stem-final /u/ or /w/ (cf. for Arabic *aḥ* 'brother', *aḥu-* 'brother of',

iḥwa, *iḥwān* ‘brothers’; also *uḥt*, plural *aḥawāt*, ‘sister’). In many words, however, the labiovelar appears with no apparent conditioning environment: *əg* ‘young of an animal’ (Arabic *ʕijl* ‘calf’, Hebrew *ʕegel*, Aramaic *ʕeglā*), *ḥ* “*ʕallāq*” ‘count’ (Arabic *ḥalaqa* ‘create; shape’, Hebrew *ḥālaq* ‘divide’—influence of past participle on **ḥəlluq* ‘numbered’ (?)). Note that the fact that the phenomenon seems limited to velars may argue in favor of /ḥ/ being a spirant in the velar region, even though it merged with the pharyngeal /ħ/.

12.3.1.4. /š/

This consonant corresponds most often to Arabic /š/, Hebrew /š/ (and hence to Epigraphic South Arabic voiceless lateral /s₂/, whose grapheme Ge’ez uses for this consonant). Out of 34 lexical items with initial <ś> in Ge’ez for which cognates are cited in both Hebrew and Arabic in Leslau 1987, about half follow this pattern; the others show various combinations of Hebrew and Arabic /s š ʃ/. If /š/ was its phonetic value, then it merged with /s/, the only pronunciation known by the pronunciation tradition. Such a merger of the palatal(-alveolar) spirant /š/ with the alveolar /s/ could possibly be connected with the emergence of a new series of palatal(-alveolar) spirants and affricates in Ethio-Semitic (e.g., in Amharic a new series /š ž č ʕ j ñ/ is created from /s z t ʈ d n/ before /i, e/, with the new /š/ moving into phonetic space vacated by the old. However, one cannot exclude the possibility of a development of /s₂/ in Ethiopic and South Arabic independent of that in Arabic and other Semitic sub-groups. Note in any case that on the one hand there is a great deal of inconsistency in the correspondence sets, while on the other, the manuscript tradition is very poor, variant writings are very common, and it is frequently difficult to determine whether the citation form of a lexical item should be given with <s> or with <ś>.

12.3.1.5. *Laryngeals* /ʔ ʕ ħ h/

As noted above, in the pronunciation tradition the only correct values for these consonants are /ʔ h/. Littmann (1918) relates anecdotally that when he read aloud a Ge’ez text before some learned clerics in Axum (thus in Tigrinya-speaking territory), giving the common Semitic values to these consonants, he was accused of pronouncing Ge’ez “like a peasant”—i.e., a naive Tigrinya speaker who had not learned the more elegant, courtly Amharicizing pronunciation. Note also that in Ge’ez, as in many varieties of Semitic, there is no phonological distinction in word-initial position between simple vocalic and glottal onset, even though the writing system has

to use a glottal-stop sign to “carry” the vowel. In the following we will not indicate an initial /ʔ/ in our transcriptions, even though rule 5 in § 12.4.2 is proof of its historical, and perhaps systematic, existence.

12.3.2. Vowels

Ge'ez has the seven-vowel system shown in Table 12-3.

Table 12-3. Ge'ez Vowels

i	ə	u
e	o	
ä		
	a	

/ä/ represents Semitic short low(-central) /a/. Length distinctions having been replaced by position-of-articulation as a primary phonological distinction in Ethio-Semitic generally, /ä/ is low central front, higher and more forward than /a/, secondarily perhaps also shorter; approximates IPA [æ].

/i u a/ are the primary high-front, high-back, and low(-central) vowels of Ge'ez. Historically they correspond to the Semitic basic long vowels /ī ū ā/. /e o/ can represent morphophonemically and etymologically earlier/underlying /äy äw/. See § 12.4.3.

/ə/ is pronounced as a high, central-back, unrounded vowel. It represents a merger of the Semitic short high vowels /i u/. (Note that loss of distinction between these two vowels is a common feature of the Damascus Arabic vowel system, among others.) This is also the unmarked epenthetic vowel used to resolve unacceptable or less preferred consonant clusters which can arise in the course of inflectional and derivational processes. As noted above, the Ethiopic writing system does not distinguish between a consonant followed by /ə/ and a word- or syllable-final consonant—and thus makes its presence in practice non-distinctive. The question of the “phonematicity” of /ə/ has in fact been explicitly posed for Amharic. For the traditional pronunciation of Ge'ez, however, a number of derivational, lexical, and phonological criteria point to the necessity of treating /ə/ as a true phonological vowel, at least on a systematic phonemic level of representation. Thus it occurs (stressed and unstressed) independently of considerations of consonant clustering in first and second person possessive suffixes (cf. paradigm in § 12.5). Likewise it is subject to the traditional pronunciation's stress-assignment rules like any other vowel. For example, the penultimate

syllable verb stress rule (see § 12.5) affects the penultimate /ə/ in *tənäggǫri* (2 f. sg. imperfective) just as it does the /ä/ in *näggǫrä* (3 m. sg. perfective). Finally, certain morphological patterns have a distinctive /ə/ which cannot be dropped, and which is independent of the rules for final consonant-cluster resolution given below: thus “conative” (“third derived form”) jussive *yəbārək* ‘may he bless’, never ***yəbārək* (contrast *bāhr* <bāḥr> ‘sea’) [here and in the following, double ** will be used for non-occurring forms, as opposed to single * for reconstructed forms]; likewise the so-called quadriliteral plural pattern *ä ... a ... ə*, as in *dānagəl*, plural of *dəngəl* ‘virgin’.

12.4. Sequential constraints

12.4.1. Syllabic structure

The preferred syllable type has a single vocalic nucleus and at most a single consonantal onset and/or coda: (C)V(C). When unacceptable or less preferred syllable types arise in morphological processes or loanwords, various strategies are invoked.

12.4.1.1. Word-initial clusters

The following epenthesis rule accounts for the resolution of the vast majority of native Ge‘ez initial consonant clusters:

$$\emptyset \rightarrow \text{ə} / \#C _ C$$

Thus the imperative, whose systematic form for the simple stem of the verbs *ngr* ‘speak’ and *lbs* ‘wear’ would be *ngər*, *lbäs*, become *nəgər*, *ləbäs*. There are very few exceptions to this rule, the only widely cited ones being *krämt* (variant *kərämt*) ‘winter, rainy season’ and *krəstos* ‘Christ’ with derivatives *krəstiyan* ‘Christian’, *krəstənnä* ‘Christianity’, *krəstun* ‘one who has become a Christian’. There is no productive rule for breaking up clusters with an initial vowel (such as exists in Arabic). There are, however, isolated lexical patterns such as *əgzi?* ‘lord’ from the root *gz?* ‘rule’ (the noun occurs in the very high frequency compound word *əgzi?äbḥer* ‘God’ [literally, ‘lord of the world’]), *əngəda* ‘stranger, guest’ from the root *ngd* ‘travel’. Note also the form of some common complementizers and conjunctions *əsmä* ‘because’, *ənzä* ‘while’, *əskä* ‘until’ (of uncertain etymology, although the first may be connected to *səm* ‘name’). Fully lexicalized loanwords apparently break up unacceptable clusters not by this rule, but by more ad hoc prothesis and epenthesis processes. Note by way of example: *sätoṭira* ‘coin’ (from Greek *statēr*), *askären* ‘box’ (from Greek *skrinion*), *askema* ‘monastic garb’ (from Greek *skhēma* ‘dress’, cf. Syriac *ʔeskēmā*).

12.4.1.2. Word-final clusters

Investigations of traditional pronunciation are in agreement that, as opposed to word-initial clusters, word-final clusters of two consonants are not broken up by an epenthetic vowel. Makonnen (1984) finds that all word-final clusters are subject to the rule

$\emptyset \rightarrow \text{ə} / \text{CC} _ \#$

and uniformly transcribes: *bahrə* 'sea', *ṭābbābtə* 'wise men' (m. pl. of *ṭābib*), *yəblə* 'he says, he will say' (irregular imperfective of *bəhlä*). Cohen (1921) and Mittwoch (1926) interpret words such as these as having word-final clusters, and usually transcribe them as such (Mittwoch adds that word-final /r, l, n/ are "syllabic" when post-consonantal, but says nothing about other clusters). Thus: *bahr*, *ṭābbābt*, *yəbl*. Cohen does, however, note a tendency to pronounce these words with a final /ə/ when they are used in isolation, and even notes this pronunciation in context when the cluster involves a /ʔ/: *ṣərʔ* 'Greece', pronounced both *ṣərə* and *ṣərʔə*. What is at stake here is of course accuracy of facts about phonetic performance, and not systematic phonological representation. All investigators are in agreement that the /ə/ in words such as *yəbarək* and *dānagəl* are part of the word's vowel pattern, and not the result of epenthesis.

12.4.1.3. Clusters of three consonants

Given what has been said above about the fully vocalic nature of /ə/, it is not clear whether the morphology as such ever confronts the speaker with clusters of more than two consonants—even though the reader of Ge'ez texts frequently has to face the problem of interpreting a sequence of three or more consonant graphemes in the sixth order. Thus the word written <qätqt> 'break!' (root *qtqt*) is to be pronounced *qätqəṭ* and not **qätəqṭ* because the imperative vowel pattern for the quadriliteral verb is *CäCCəC*. Note in this context that the /t/ suffix which frequently occurs in plurals and in feminines regularly clusters with the preceding consonant, and never occurs in the shape -əṭ. Thus *nəgəst* 'queen' (= *nəgəs* + *t*, not *nəgs* + *t*; cf. *nəgus* 'king', also *nägäst* 'kings').

12.4.1.4. Gemination

Gemination is a widely employed inflectional and derivational process in Ge'ez, as in the rest of Ethio-Semitic. To give a few examples, it is associated with lexical classes (verbs of lexical class II have a geminated middle radical in all inflected forms: *fäṣṣāmä* 'he completed' vs. *nägärä* 'he spoke'), tense

inflection (gemination of middle radical in imperfective: *yəṇāggər* ‘he speaks, will speak’), nominal derivation (*fällas* ‘sojourner’ from *fäläsä* ‘sojourn’, *hāqqal* ‘countryman, farmer’ from *hāql* ‘field’), plural formation (*tābbābt* ‘wise men’ from *tābib* ‘wise man’), suffixation in certain environments (especially involving the “short” vowels /ä ə/: *nägäränni* ‘he spoke to me’ vs. *nägärūni* ‘they spoke to me’, *nägärätto* < *nägärät+ä-hu* ‘she spoke to him’ vs. *nägärätänni* ‘she spoke to me’, *yəngərro* ‘let him speak to him’ < *yəngər* ‘let him speak’, *nägärkänni* ‘you [f. sg.] spoke to me’ < *nägärki* [cf. § 12.4.5]). In the traditional pronunciation, all consonants can geminate except the laryngeals. In general, gemination seems to be limited to vowel-pattern environments in which the introduction of gemination will not give rise to problematic consonant clusters. (Mittwoch 1926, however, consistently notes some gemination patterns of which this is not the case: *nägärkä* ‘you spoke’ but *nägärkkänni* ‘you spoke to me’, and regularly *nägärkkāmu* ‘you (m. pl.) spoke’, as opposed to others’ *nägärkāmu*. These cases need more investigation, possibly in connection with the problems raised in § 12.5.)

12.4.2. Laryngeal effects

A series of rules affect the vowels /ä ə/ in the vicinity of laryngeals. (In the following, H stands for the set of laryngeal consonants /ʔ ʕ ħ h/, and H_{*n*} is a laryngeal in the *n*th position in a root.) Although the immediate phonological environment and output of these rules is rather straightforward, the diachronic and synchronic interaction of these rules with one another and with the morphology results in a complex set of constraints on words, especially verbs, containing laryngeals. There are various ways in which these constraints could be formulated in detail, depending on theoretical framework, and analysis of morphology. The following is simply a descriptive statement of the phenomena, with some illustrative examples.

1. ə → ä / __ H ä
2. ä → ə / __ H [V, high] ([V, high] = /i ə u/)
3. ä → ə / __ H_{fin} V
4. ä → a / __ H {C, #}
5. ä → a / H __

Examples (where the affected segment is underlined, with a corresponding non-laryngeal form in parentheses):

1. a) *yähärrəs* ‘he plows’ (*yəṇāggər* ‘he speaks’)
- b) *lähäkk* ‘send!’ (*ləbäs* ‘wear!’)
- c) *yəsämməhäkkä* ‘he hears you’ (*yəṇāggərhäkkä* ‘he speaks to you’)

- d) *abagəfä* 'sheep' [acc. pl. of *bägf*] (*awalədä* 'daughters' [acc. pl. of *wälätt*])
2. a) *rəhib* 'wide' (*täbib* 'wise')
- b) *yəməhər* 'he is merciful' (*yənəggər*)
- c) *yəkəfu* [root *kfw*] 'he pours' (*yətəllu* [root *tlw*] 'he follows')
- d) *ləh(ə)qä* 'he grew up' (*läbsä* < **läbəsä* < **läbisä* 'he wore')
- e) *tä?əhzä* 'it was held' (*tänägrä* < **tänägirä* 'it was said')
- f) *yəsməfu* 'may they hear' (*yəlbäsu* 'may they wear')
3. a) *sämfiä* < **säməfiä* < **sämiä* 'he heard' (*läbsä* < **läbəsä* < **läbisä* 'he wore')
- b) *asməfiä* 'he caused to hear' (*albäsä* 'he clothed')
- c) *nässəhā* 'he repented' (*fässämä* 'he finished')
- d) *baləhā* 'he liberated' (*barākā* 'he blessed')
- e) *bəhbəhā* 'it roared (water)' [root *bhbh*] (*qätqätä* 'it burst' [root *qtqt*])
4. a) *yəsməf* 'may he hear' (*yəlbäs* 'may he wear')
- b) *säməfku* 'I heard' (*läbäsku* 'I wore')
- c) *mə?käl* 'midst' [root *?kl*] (*mängäd* 'road' [root *ngd*])
5. a) *əmnä* 'he believed' (*läbsä* 'he wore')
- b) *haqäfä* 'he embraced' (*nägärä* 'he spoke')
- c) *lä?akä* 'he sent' (*nägärä* 'he spoke')

Rules 1 and 2 together impose what amounts to a vowel harmony constraint across laryngeals (a short vowel before a laryngeal must have the same height as a vowel which follows). However, while rule 1 applies to H_1 and H_2 verbs, as shown by 1a and 1b, 1c–d show that the constraint does not apply to H_3 triconsonantal roots (which are governed by rule 3). Examples 2a–e show that rule 2 applies to H_2 , and 2f shows that it applies (redundantly with rule 3) to H_3 , 2e shows that it does not apply to H_1 . 2d shows that H_2 verbs tend to be of the so-called "intransitive" *läbsä* type, and behave as though there were a /ə/ between the second and third root consonant. This prevents rule 4 from operating on these verbs, which would otherwise become ***lahqä* (note that example 5c shows that there are a few *nägärä* type H_2 verbs. Example 3a shows that base-form H_3 triradicals are also formally assimilated to the *läbsä* class; but as examples 3b–d show, this rule also gives rise to forms which are not paralleled in non-laryngeal verbs. 3e shows that this rule also applies to the *fourth* consonant of quadriradical roots, and that rule 3 is properly formulated in terms of *root-final* laryngeal (H_{fin}), and not just H_3 . The examples for rule 4 (plus 3e) show that it applies to H_1 and H_{fin} ,

but to H₂ only in quadriradicals. Rule 5 “lengthens” /ä/ after H (something not always reflected in the writing). A corollary of rule 5, plus the fact that graphic and systematic word-initial /ʔV/ is pronounced as word-initial /V/ in the traditional pronunciation, is that no word can begin with the vowel /ä/ (cf. 5a).

12.4.3. *Glide effects*

Vowels and glides interact in a number of assimilatory and dissimilatory processes. To begin with labiovelars (where K = /k q g ɣ/, K^w the corresponding labiovelars, and ~ is to be read ‘alternates with’), note the neutralization of the distinction between a labiovelar followed by /ə ä/ and a velar followed by /u o/:

K^wə ~ Ku (q^wəl ~ qul ‘bunch of grapes’)

K^wä ~ Ko (q^wätit ~ qotit ‘slender’)

Correspondingly, as already noted, a labiovelar cannot be followed by a rounded vowel:

*K^wu, o

The following rules handle short-vowel glide sequences in syllable final position:

äw → o (fätäwkä ~ fätokä ‘you loved’)

äy → e (sätäykä [sätekä rare] ‘you drank’)

əw → u (*yəfättəw > yəfättu ‘he loves’)

əy → i (*yəsättəy > yəsätti ‘he drinks’)

Of these rules, the first is optional, the second is rare, while the third and fourth are obligatory. The opposite order of glide and /ə/, at least post-consonantly, is operative in:

wə → u / C __ (cf. labiovelar rules)

yə → i / C __ (akyəst ~ akist, plural of käysi ‘serpent, dragon’;
cf. agbərt plural of gäbr ‘slave’)

There is also, in a glide context, a reduction in the reverse direction:

u → ə / ww __ (məwwut ~ məwwət ‘dead’, participle of mwt ‘die’;
cf. fəṣṣum ‘finished’ < fṣm ‘finish’)

i → ë / yy __ (qäyyih ~ qäyyəḥ ‘red’, adjective from root qyḥ; cf.
ḥäddis ‘new’, adjective from root ḥds)

Sequences of /u i/ and a following vowel are resolved by a process that in some sense is also a reverse of the glide+ə/ rules. (The only relevant cases seem to be those of the third person object suffix pronouns -o, -a, -omu, -on < hu, ha, homu, hon ‘him, her, them (m.), them (f.).’)

$u \rightarrow \text{ə}w(w) / _ V$ ($qātālu+o > qātāləwwo$ 'they killed him', $qātālu+a > qātāləwwa$ 'they killed her')

$i \rightarrow \text{ə}y(y) / _ V$ ($qātālki+o > qātālkəyyo$ 'you (f.) killed him', $qātālki+a > qātālkəyya$ 'you (f.) killed her')

The preceding two statements combine a glide formation and reduction process into one step. This glide formation occurs independently between the proclitic negation marker *i* 'not' and verbal forms beginning with a vowel. Thus: $i+asməfä > iyasməfä$ 'he did not cause to hear', $i+əsämməf > iyəsämməf$ 'I do not hear'. Some combination of glide formation and glide+vowel reduction is probably involved in the accusative form of nouns ending in /i/. Normal accusative is marked by suffixation of /ä/: *nəgus* 'king [nom.]' *nəgusä* 'king [acc.]'. Thus we have: *bəʔsi* 'man [nom.]' and $bəʔsi+ä > bəʔsiyä > bəʔsi(\text{ə})yä > bəʔse$ 'man [acc.]'.

12.4.4. Consonant assimilation

The only productive consonantal assimilation processes involve consonant sequences which arise across morpheme boundaries in the course of affixation. There are only three consonants involved in affixation in Ge'ez, /t n k/, and of these, only /t k/ are involved in assimilation. /k/, which occurs in suffixes marking first and second person subject in the verb, assimilates to a preceding (i.e., stem-final) velar stop:

$k \rightarrow \{g, q\} / \{g, q\} _$ ($ḥädäg+kä > ḥädäggä$ 'you left'; $säräq+ku > säräqqu$ 'I stole')

/t/ occurs as a prefix marking passive-reflexive derived stems in the imperfective tense (the shape of the prefix in the perfective is *tä-*). It assimilates fully to a following dental stop or sibilant, /t ʈ d s ʃ z ś ʈ/ = D:

$t \rightarrow D / _ D$ ($yə+t+sämmäʃ > yəssämmäʃ$ 'he is heard'; cf. $yəṯqättäl$ 'he is killed')

As a suffix /t/ marks feminine (and also plural) forms of nouns and participles. Here it assimilates to a preceding dental stop:

$t \rightarrow \{d, ʈ\} / \{d, ʈ\} _$ ($kəbud+t > kəbədd$ 'heavy [fem.]; $qəʔqut+t > qəʔqəʔt$ 'crushed [fem.]')

Note, however, idiosyncratically 'one [fem.]' *aḥatti* < *aḥad+ti*, 'daughter' *wälätt* < *wäläd+t*.

12.4.5. Vowel reduction

There are two "shortening" rules (recall that /i a u/ come from common Semitic /ī ā ū/, while /ə/ represents a merger of Semitic /i u/ and /ä/ is Semitic

/a/). One is a dissimilation rule which reduces /i/ which is part of a feminine or plural suffix before a suffix containing an /i/:

$i \rightarrow \text{ə} / + (C) \text{ ___ } + Ci$

Examples: *hāb+i+ni* > *hābənni* ‘give [f. sg.] me!’ (vs. *hābina* ‘give [f. sg.] us!’); *afras+i+ki* > *afrasəki* ‘your [f. sg.] horses’ (vs. *afrasihi* ‘his horses’); *nāgār+ki+ni* > *nāgārəkənni* ‘you [f. sg.] spoke to me’ (vs. *nāgārkinä* ‘you [f. sg.] spoke to us’); contrast *qātaliki* ‘your [f. sg.] killer’, where the *-i* is part of the ‘agentive’ vowel pattern *CäCaCi*. The other rule “shortens” /u i/ (and /a/ optionally) in syllables closed by the feminine suffix *-t*.

$u, i \rightarrow \text{ə} / \text{ ___ } C+t (kəbur+t > kəbərt \text{ ‘honored [f.]’; } ləhiq+t > ləhəqt \text{ ‘grown up [f.]’})$

$a \rightarrow \text{ä} / \text{ ___ } C+t (sānnayt \sim sännäyt \text{ ‘beautiful [f.]’ vs. } sännay \text{ ‘beautiful [m.]’})$

12.5. Stress

All modern observers agree that the traditional pronunciation assigns stress to Ge‘ez words in context, although there is a certain amount of difference on the details of this stress assignment. Ethiopian experts on Ge‘ez themselves can point out minimal pairs such as: *yənāggǝra* ‘they [f. pl.] speak’ vs. *yənāggǝrā* ‘he speaks to her’ (cf. Mittwoch 1926: 32). The following generalizations seem to be supported by the formulations and transcriptions of most modern investigations:

1. Verbs are generally stressed on the penult: *nāgārā*, *tənāggǝri* ‘he spoke, you (f. sg.) speak’

But the following have final stress:

- a) perfective, 2 f. pl. subject: *nāgārəkən* ‘you spoke’
- b) verbs with 2 f. pl. object: *yənāggǝrā(k)kən*, *nāgārā(k)kən* ‘he speaks/spoke to you’
- c) verbs with 3 sg. and 3 f. pl. object: *nāgārō*, *nāgārón* ‘he spoke to him/them’

Except: jussive, imperative: *yəngǝrro* ‘may he speak to him’ forms with /ə/ < /u, i/: *nāgārǝwwō* ‘they spoke to him’

2. Nouns and pronouns have final stress: *nəgús* ‘king’

Except: When followed by the *-ä* accusative marker: *nəgúsä* ‘king [acc.]’

3. Personal pronouns and possessive suffixes follow special patterns. The most common forms are stressed as follows:

	Sg	Pl	Sg	Pl
1	ánä	nə́hñä	-ə́yǎ	-ə́nǎ
2m	ántä	antə́mu	-ə́kǎ	-əkə́mu
2f	ánti	antən	-ə́ki	-əkən
3m	wə́ʔətu	ə́muntú	-(h)ú, -ó	-(h)ómu
3f	yə́ʔəti	ə́mantú	-(h)á	-(h)ón

It cannot, however, be overemphasized that no complete analysis of stress patterns have been made for *any* Ethio-Semitic language, and that much remains to be done in this area. Cohen's intuition (1921: 230, 258²) was that in Ethio-Semitic factors of intensity, syllable length, gemination (and conceivably pitch, Kolmodin 1910: 250) may be interrelated in a complex metrical pattern whose key has not yet been discovered. We still need to find out whether he may have been right.

Bibliography

- Cohen, Marcel. 1921. "La prononciation traditionnelle du guèze (éthiopien classique)." *Journal Asiatique* ser. 11, vol. 17: 217–69.
- Dillmann, August. 1857. *Grammatik der äthiopischen Sprache*. Leipzig: Tauchnitz.
- . 1907. *Ethiopic Grammar*, 2nd ed., enlarged and improved, by Carl Bezold (1899). Trans. James Crichton. London: Williams & Norgate.
- Kolmodin, Johannes. 1910. "Meine studienreise in Abessinien 1908–1910." *Le Monde Oriental* 4: 229–55.
- König, Eduard. 1877. *Neue Studien über Schrift, Aussprache und allgemeine Formenlehre des Aethiopischen*. Leipzig.
- Lambdin, Thomas. 1978. *Introduction to Classical Ethiopic (Ge'ez)*. Missoula, Mont.: Scholars Press.
- Leslau, Wolf. 1987. *Comparative Dictionary of Ge'ez*. Wiesbaden: Harrassowitz.
- Littmann, Enno. 1918. "Ge'ez-Studien I, II, III." *Nachrichten der K. Gesellschaft der Wissenschaften zu Göttingen*, Philologisch-historische Klasse, 627–702; 318–39.
- Ludolf, Hiob. 1661. *Grammatica aethiopica*. London: Thomas Roycroft.
- Makonnen Argaw. 1984. *Matériaux pour l'étude de la prononciation traditionnelle du guèze*. Paris: Editions Recherche sur les Civilisations.
- Mittwoch, Eugen. 1926. *Die traditionelle Aussprache des äthiopischen* (Abessinische Studien 1). Berlin: de Gruyter.

Trumpp, Ernst. 1874. "Ueber den Accent im Aethiopischen." *Zeitschrift der deutschen morgenländischen Gesellschaft* 28: 515–61.

Ullendorff, Edward. 1955. *The Semitic Languages of Ethiopia: A Comparative Phonology*. London: Taylor's (Foreign) Press.

Arabic Phonology

Alan S. Kaye

California State University, Fullerton

13.1. Arabic and its relationship to the other Semitic languages

The Semitic languages (part of a larger phylum generally called Afroasiatic [Afro-Asiatic], Afrasian, or Hamito-Semitic) are usually classified into East Semitic, consisting of Akkadian with its two major dialects, Assyrian and Babylonian, and West Semitic. West Semitic, in turn, consists of (1) Northwest Semitic with Hebrew, Aramaic (a complicated variety of dialects and very distinct languages by any usual definition of this term, such as Syriac, the language of the Peshitta translation of the Bible), Amorite, and Phoenician-Punic (most scholars also include here Ugaritic but not Eblaite [the latter of which seems most closely related, in all probability, to Akkadian], two ancient cuneiform languages found in Syria in the 1920s and 1970s, respectively); and (2) Southwest Semitic comprising Arabic (Epigraphic, or Old, and Modern South Arabian), and Ethiopic (a conglomerate of many different languages and dialects). Arabic is sometimes referred to as North Arabic to differentiate it from South Arabian. Despite the similarity in name, Epigraphic South Arabian, also called Epigraphic South Arabic (ESA), is not any closer to Arabic than to some other Semitic languages. ESA is known in four major dialects (Sabaeen, Qatabanian, Minaean, and Ḥaḍramatic), which are closely related to the Ethio-Semitic languages. This is quite understandable since South Arabian speakers came to Ethiopia from today's Yemen bringing their language and writing system with them. The modern South Arabian languages, such as Soqotri and Shxauri (the latter is also known as Šheri or Jibbāli), are not so close to (North) Arabic. They are mutually unintelligible with it, even though they have borrowed many Arabic lexemes due to the influence of Islam and the languages-in-contact situation.

AUTHOR'S NOTE: I wish to record my gratitude to the following colleagues, who contributed many useful comments on a preliminary version of this chapter: Peter T. Daniels, Stuart Davis, Robert Hoberman, Jeffrey Heath, Gary A. Rendsburg, and Stanislav Segert. Needless to say, the usual disclaimers apply, especially since I have sometimes not followed the advice offered.

Table 13-1. Comparison of the Major Semitic Languages

	Akkadian	Hebrew	Aramaic	ESA- Ethiopic	CA	
1.	<i>kalbu</i>	<i>kēlev</i>	<i>kalbā</i>	<i>kalb</i>	<i>kalbun</i>	‘dog’
2.	<i>šūru</i>	<i>šōr</i>	<i>tawrā</i>	<i>sōr</i>	<i>θawrun</i>	‘bull’
3.	<i>rēšu</i>	<i>rōš</i>	<i>rēšā</i>	<i>rēʔs</i>	<i>raʔsun</i>	‘head’
4.	<i>ʔēnu</i>	<i>ʔayin</i>	<i>ʔaynā</i>	<i>ʔayn</i>	<i>ʔaynun</i>	‘eye’
5.	<i>ʔuznu</i>	<i>ʔozen</i>	<i>ʔuðnā</i>	<i>ʔezn</i>	<i>ʔuð(u)nun</i>	‘ear’
6.	<i>lišānu</i>	<i>lošōn</i>	<i>liššānā</i>	<i>læssān</i>	<i>lisānun</i>	‘tongue’
7.	<i>ʔūmu</i>	<i>yōm</i>	<i>yawmā</i>	<i>yōm</i>	<i>yawmun</i>	‘day’
8.	<i>ʔeqlu</i>	<i>heleq^a</i>	<i>haqlā</i>	<i>haql</i>	<i>haqlun</i>	‘field’
9.	<i>šalāmu</i>	<i>šālōm</i>	<i>šlāmā</i>	<i>salām</i>	<i>salāmun</i>	‘peace’
10.	<i>napištu</i>	<i>nefeš</i>	<i>nawšā</i>	<i>nafs</i>	<i>nafsun</i>	‘soul’

a. Not listed by Leslau (1987: 240) but listed by Köhler-Baumgartner (1958: 306).

Some Semitic languages are compared in Table 13-1, which represents data from the ancient Semitic languages: Akkadian (Assyro-Babylonian), Hebrew (Classical or Tiberian), Aramaic (Syriac), reconstructed ESA since the vowels were not written, Classical Ethiopic (Ge‘ez), and Classical Arabic (CA). ESA and CA are the most conservative of all the Semitic languages, phonologically speaking. Akkadian, on the other hand, the oldest attested Semitic language, has lost the pharyngeals, among other noteworthy developments. Most modern colloquial Arabic dialects use descendants not very different from the ten sample words listed above. Thus *kalb* ‘dog’ and *rās* ‘head’ still survive in the great majority of Arabic vernaculars spoken today.

The Semitic family of languages operates on the root + vocalic pattern principle, with Arabic being the most systematic representative of the famous triconsonantal system and the one usually used by linguists as typically illustrative of Semitic. The root consists of consonants only (usually three), while the vowels often express the grammatical categories. Thus, the standard root (used by many Semitists) QTL has something to do with ‘killing’, such as CA *qatl* ‘murder’, *qatala* ‘he killed’, *maqtūl* or *qatīl* ‘killed one (m.)’, *qātil* ‘murderer, killer (m.)’, *qītāl* ‘battle’, *qātala* ‘he fought with’, etc. The three radicals (or root letters) are not pronounceable by themselves, and often roots do not share a single semantic sphere. For example, *ʕaliqa* means

‘to hang’ whereas *ʕallaqa* is ‘to comment’. Some vocalic patterns, as the latter form illustrates, also operate with gemination of one of the radicals; e.g., *qattala* means ‘he massacred’ with the gemination of the second radical (C²) of the root (the second consonant of the root is geminated in Form II of the verb, the so-called **intensive**).

The consonantal root and vocalic patterning system was probably at least partially responsible for the evolution of the Arabic script, developed out of Aramaic (Nabatean) via the “Sinaitic” script, which was molded from the West Semitic signary. ESA and Phoenician do not express any vowels at all. Thus, the Phoenician graphemes <QTL> may mean ‘he killed’ (= CA *qatala*, Hebrew *qāṭal*) or ‘she killed’ (= CA *qatalat*, Hebrew *qāṭalā*), even though their pronunciations were distinct. The West Semitic alphabets (Hebrew, Aramaic, and Arabic) have all made use of *matres lectionis* (‘mothers of reading’), a device utilizing the morphophonemically weak consonants, viz., the glottal stop and approximants (*ʔ*, *h*, *w*, *y*), to express long vowels, although there were and still are differences of usage among the various Semitic languages. The short vowels are indicated by special markings of rather late origin and are not normally used in the modern written language (Modern Standard Arabic); for CA these markings include <a>, <i>, and <u>, *shadda* or *tašdīd* (the gemination grapheme) and other graphemes for the correct recitation of texts, i.e., the *tajwīd* of the Holy Qur’an (see especially Nelson 1985: 14–31). This classical–modern situation is similar in some ways to Classical (Biblical) and Modern (Israeli) Hebrew and their intricate interrelationships.

Arabic has survived as a spoken language, whereas Akkadian and the Canaanite dialects have died out with the exception of Hebrew, if indeed Hebrew is designated as Canaanite. Aramaic (the mother tongue of Jesus) still survives as many different spoken languages in Iraq, Iran, Turkey, and Syria (and of course, countries such as Israel and the U.S.A., to which those Middle Eastern peoples have immigrated). It is known today as either Neo-Aramaic/Neo-Syriac or modern Assyrian (rarely also as Chaldean). Many Aramaic speakers, it should be noted, also speak an Arabic dialect. The various Neo-Aramaic languages have, taken all together, approximately a quarter million speakers. (In many ways, Aramaic has developed into many Neo-Aramaic languages in a fashion similar to Latin’s developing into the Romance languages.) Classical Ethiopic or Ge’ez has not survived except as the liturgical language of the Ethiopic Church (paralleling the situation of Coptic in Egypt), and there are many modern, spoken Ethio-Semitic

languages, such as the national language of Ethiopia, Amharic, or the two northern Ethiopian languages, Tigre and Tigrinya, to which Ge'ez is closely related in a northern Ethiopic sub-branch. There are many Arabic loanwords in the Ethio-Semitic languages (see Leslau 1990).

13.2. Arabic dialectology

Arabic is the Semitic or indeed the Afroasiatic language with by far the greatest number of speakers, probably now in excess of 200 million. It is the major language throughout the Arab world—Egypt, the Sudan, Libya, the closely knit North African countries (usually referred to as the Maghrib, which includes Tunisia, Morocco, and Algeria), Mauretania, Saudi Arabia, Iraq, Lebanon, Syria, Jordan, the Gulf states, etc. It is even the major language of a non-Arab country, the Republic of Chad in central Africa (i.e., more Chadians speak Arabic as their mother tongue or as a second language than any other single language).

As a minority language, Arabic is spoken and widely studied in other nations such as Nigeria, Iran, Afghanistan, Israel (where it is an official language), and some Inner Asian countries, although Uzbeki Arabic may now be completely extinct. Furthermore, it is in wide use throughout the Muslim world as a second language and as a learned, liturgical language (e.g., in Pakistan, India, Tanzania, or Indonesia). Indeed, among orthodox Muslims, Arabic is *luġhat al-malāʾika* 'the language of the angels', and the language *par excellence* in the world since Allah himself is believed to have revealed his Holy Book, the Qur'an, in the Arabic language. One can also readily comprehend that Arabs are very proud of their (most beautiful) language by considering that there is even an Arabic verb *ʔaṣraba* 'to speak clearly and eloquently', which has come to mean 'to express' in general, a derivative of the root ʕRB, also occurring in the word *alʕarabiyyah* 'the Arabic language', or *lisān ʕarabī* 'the Arabic language' in the Qur'an.

There are, from a purely descriptive point of view, many recognizably distinct, major Arabic dialects. The peripheral Arabic dialects are, in fact, so radically different from those of the main Middle Eastern core (in which there is somewhat of a cultural solidarity) that they are better referred to as separate languages, by any satisfactory definition of what a language is (see Kaye 1994: 47). Certainly mutual unintelligibility is an overriding factor here. (I am familiar with the old adage that the major difference between a language and a dialect is that a language is a dialect with an army and a

navy.) One of these deserves very special mention. Maltese is unquestionably a dialect of Arabic from the historical point of view. This historicity has produced the rather awkward designation “Maltese Arabic.” Due to its isolation from the rest of the Arab world where we have noted a cultural uniformity (see above), Maltese has developed into a new Semitic language in its own right (a similar, but slightly weaker, argument could also be made for Cypriot Maronite Arabic, now probably extinct on Cyprus). The two major reasons for not regarding Maltese as a dialect of Arabic synchronically are (1) it does not have diglossia, i.e., it does not have CA as a “high” level of language (more on this important topic later); and (2) it is written in the Latin script. Both of the aforementioned points taken together would certainly make it unique as a dialect of Arabic. Furthermore, the Maltese do not consider themselves Arabs, nor do they consider their variety of English pronunciation to be part of any so-called Arabic “accent” in the widest possible sense of this term.

13.3. Arabic as Central Semitic

According to Robert Hetzron’s innovative classification of the Semitic languages (1972: 16), which is not universally accepted by Semitists, Arabic shares traits of both South Semitic and Northwest Semitic. Arabic preserves Proto-Semitic (PS) phonology almost perfectly (ESA is even a bit more conservative), except for PS **p > f* and **ś > š* (note that PS **ś* and **š* develop into Arabic *s*). But it also shares features with Hebrew, Ugaritic, and Aramaic, such as the masculine plural suffix *-īna/īma* and the internal passive, e.g., Arabic *qatala* ‘he killed’ vs. *qutla* ‘he was killed’, Aramaic *qəṭīl* ‘killed’ and Arabic *qatīl* ‘killed one’; furthermore consider Hebrew *hīlbīš* ‘he dressed someone’ vs. *hulbaš* ‘he was dressed (by someone)’.

The morphology of the definite article in Hebrew (*ha-* < **han* + gemination of the following consonant, where possible) and Arabic (*ʔal-*, cognate with **han* and Ugaritic /*hn-*/ in the demonstrative pronouns *hnd* and *hnk*, which assimilates before dentals, sibilants, or sonorants, producing a geminate) also points to a common origin. The Hebrew *ha-*, in fact, shows up, in my view, in the Arabic demonstratives: *hāḏā* ‘this (m.sg.)’, *hāḏihi* (f.), and *hāʔulāʔi* (pl., all genders). Even the broken plurals of Arabic, in which one can see the vocalic change in the stem (cf. Arabic *kilāb* ‘dogs’, pl. of *kalb*), may be compared with Hebrew *segholate* (viz., C¹εC²εC³) plurals, such as *kālābīm* ‘dogs’ (cf. Hebrew singular *keleb* < **kalb* + masculine plural *-īm*).

There are some other striking morphological affinities of Arabic with Hebrew, such as the ancient dialectal Arabic relative particle $\delta\bar{u}$ = Biblical Hebrew $z\bar{u}$, while the Western form $\delta\bar{i}$ occurred in Arabic $\text{?alla}\delta\bar{i}$ ‘who (m.sg.)’ and Aramaic $d\bar{i}$. In all likelihood, the form with u represents the nominative while the form with i retained the oblique (i.e., genitive and accusative) marking. Some Eastern dialects also reflected Barth’s Law; i.e., they had i as the imperfect preformative vowel with a of the imperfect stem, as in the Canaanite dialects.

13.4. Consonants

The consonantal segments of a fairly typical educated pronunciation of Modern Standard Arabic (MSA), identical in most respects to CA, are included in the consonant chart in Table 13-2. The symbols are modified from the International Phonetic Alphabet (IPA), as is the custom of linguists specializing in Arabic and other Semitic languages.

Table 13-2. Modern Standard Arabic Consonants

stops	b		t d	ṭ ḍ		k	q		ʔ
affricates						ʧ			
fricatives		f	θ ð	s z	ṣ ḏ (ẓ)	ʃ	x ɣ	ħ ʕ	h
nasals	m			n					
liquids				l r	ḷ				
approximants	w					y			

The Arabic alphabet provides an accurate depiction of the phonological facts of the standard language. However, in many dialects there are some pronunciations that differ from those presented in Table 13-2. For instance, CA /q/ is voiced in many dialects, both ancient and modern, i.e., [g] (especially among various Bedouins, and this is probably indicative of its original Arabic pronunciation). Moreover, the CA ?im (the name of the letter represented by the grapheme <ʔ>) corresponds to many pronunciations, such as [j], [ɟ], [dy], [gy], [g], [y], or [ʒ] stemming from PS */g/. Cairo Arabic [g] reflects a complicated chain of developments (see Kaye 1972) from Proto-Colloquial Arabic */ʒ/ (as preserved, say, by Damascus Arabic).

Every consonant in CA, MSA, and most Arabic dialects may be geminated, unlike Hebrew, for example, which cannot geminate the so-called “gutturals” ʔ, ʕ, h, ħ, and r—the pharyngeals and laryngeals plus r (which

may have been uvular as it is mostly in Modern Israeli Hebrew, although the acrolectal pronunciation is still rolled).

CA does not have a /p/, but standard pronunciations tend to devoice a /b/ before a voiceless consonant, e.g., /ħabs/—[ħaps] ‘imprisonment’ or /ħibs/—[ħips] ‘dam’. Some modern Arabic dialects (notably in Iraq) have both /p/ and /p/ (emphatic); however, the great majority of Arabic speakers produce English /p/ as /b/ through interference modification (one Arab asks another, “Which Bombay are you going on vacation to? Bombay, India, or Bombay—Pompeii—Italy?”). Incidentally, Persian, Urdu, and other languages which have /p/ have taken the grapheme for /b/ = ب and made پ by placing three dots underneath its basic configuration to = <p> (no confusion with <t> or <θ>). This grapheme, in turn, has been reborrowed by some Iraqi Arabs, who use it both in native developments as well as in Persian loanwords.

CA does not have a /v/, but phonetically a [v] often occurs through regressive assimilation as in /ħifð/—[ħivð] ‘memory’. /n/ also assimilates regressively, i.e. *nb* > *mb* (/mimbar/ منبر ‘mimbar’ and *ng* > *ŋk* as in /bank/ > [baŋk] ‘bank’).

13.4.1. The term “emphasis” as applied to Arabic

The *emphatic* consonants, often misleadingly called velarized or pharyngealized or pharyngealized-velarized, are depicted with a dot underneath the consonant. Emphatic pronunciation has little or nothing at all to do with velarization and nothing to do with what is involved in the articulation of [ħ] and [ʕ]. Perhaps nowhere else in Arabic linguistic literature is there more controversy than in this area of the emphatics—how they are to be described and how they function in a dialect and cross-dialectally. The vowels around an emphatic consonant tend to become lower, retracted, or more centralized than those around corresponding non-emphatics. The very back consonants, /q/ and sometimes /x/, /ɣ/, and /r/, have a similar effect on vowels, which is why the vowel allophonics of Arabic are much more cumbersome and intricate than the consonant allophonics. In some dialects, e.g., Cairene, vowels are lowered after /ħ/ and /ʕ/ as well: /ħub/ ‘love’ = [ħob], and /ʕumar/ ‘Omar’ = [ʕomar]. Before or after /ħ/ and /ʕ/, the short /a/ = [ɑ] rather than the common [æ] (see Harrell 1957: 50–51). /r/ also triggers the back /ɑ/ as in [maʃrubɑ:t] ‘drinks’, and there are minimal pairs indicating a phonemic difference between /r/ and /r/: *ħarri* ‘my land’ vs. *bærri* ‘pertaining to land’, *ħaʔarri* ‘my cows’ vs. *bæʔæri* ‘bovine’, *gaʔri* ‘my neighbor’ vs. *gæ:ri* ‘running’, and *ra:mi* ‘Rami (p.n.m.)’ vs. *ræ:mi* ‘throwing’ (Harrell 1957: 72–73).

In Old Arabic, the primary emphatics were most likely voiced, i.e., the /ḍ/ < [zʌ] (lateralized) IPA [ḏ], or [zʌ], the /ṭ/ < /ḍ/, the /ḏ/ or /zʌ/ < /ḏ/, and the /ṣ/ < /zʌ/. Some linguists have argued that the emphatics in Arabic and other Semitic languages were originally glottalized, as preserved in Ethio-Semitic. My own view is that glottalization in the latter case should be regarded as due to Cushitic substratum.

13.4.2. *The debate about the nature of the emphatics*

Most dialects of Arabic spoken today contain emphatic versus non-emphatic contrasts (referred to in Arabic as *ʔiṭbāq*, *tafxīm*, or *mufaxxama*). Thus, a word such as CA *ṣayf* ‘summer’ contrasts with its non-emphatic counterpart *sayf* ‘sword’. The usual phonological explanation for this feature (see Gairdner 1925) lies in the opposition of the emphatic /ṣ/ and the non-emphatic or plain /s/. In a classic article on the subject, Jakobson (1957) attributed some of the phonetic characteristics of emphasis to a contraction of the upper pharynx (usually called pharyngealization). Velarization refers in essence to a decrease in the back orifice of the buccal cavity serving as a resonator. (The IPA, unfortunately in my view, indicated both features with a wavy bar, actually a tilde, through the particular consonant, i.e., [ʕ]; for [ṣ]; however, the revised IPA [1989] noted velarized [tʰ] vs. pharyngealized [tʰ̠], whereas [ṭ] is either velarized or pharyngealized.) In addition to the above, emphasis often involves some degree of protrusion and rounding of the lips. Cf. Nigrian Arabic *amm̥* ‘mother’ vs. *amm* ‘paternal uncle’ (Owens 1993: 25). This explains why some emphatics are often perceived as labials by non-Arabs. Any phoneme in the language which is articulated with the aforementioned phonetic features is perceived by native speakers as being emphatic (although even native speakers on rare occasions refer to certain consonants as being only “slightly” emphatic, an impressionistic phonetic designation).

The description of emphasis in terms of acoustic phonetics is perhaps easier to comprehend. The emphatic phonemes are lower in pitch; i.e., the sound spectrograph has revealed that they have higher energy in the lower frequency regions. This is why Jakobson (1957) proposed to label this particular feature [+flat] (which he also used for both labials and labialized phonemes).

Depending on the analyst’s point of view, vowels may also be emphatic or, at least, influenced (via assimilation) by adjacent or nearby emphatic consonants. This means that emphatic vowels are consonantly conditioned allophones and can be predicted from the surrounding emphatic consonants.

Though emphasis has been studied from a phonetic standpoint (x-ray studies go back more than half a century), the various phonological perspectives as to its role and function in the language in terms of a more abstract or underlying (morphophonemic) level remain highly controversial. One of the main reasons for this is that emphasis is never an articulatory phonetic feature of only one phoneme. The word /ʃayf/ 'summer' (cited above) is really [ʃayf]. A word such as /tīn/ 'mud' consists of a [t] followed by [ī] and [n], i.e., [tī:n] or [t̤i:n]. Emphasis always affects two or more segments, a fact not represented in the orthography of CA. For this reason, an analysis of emphasis as a monovalent distinctive feature (present or absent in a single phoneme) is unsatisfactory.

According to a modern binary viewpoint, Arabic consonants are either [+emphatic] or [-emphatic]. All emphatic vowels would be emphatic by virtue of the influence of these phonologically emphatic consonants. The biggest problem with this type of analysis is that it immediately doubles the phonemic inventory of Arabic consonants. Another oddity is that one would have terribly cumbersome morphological (allomorphic) representations, i.e., the *-t* marking the first person singular perfect has an allomorph *-t̤* after an emphatic consonant, and so on. Consider, for example, *katab + t* 'I wrote' but *ḍaṛaḇ + t̤* 'I beat/struck'.

Still other phonologists have suggested that all vowels exist in emphatic and non-emphatic pairs. This is tantamount to stating that emphasis is a redundant feature in terms of consonants. The problems with this way of looking at things are precisely the reverse of the vocalic redundancy solution. An Arabic morpheme consists (often) of a root plus a vocalic pattern. With the "normal" triconsonantal root, the discontinuous *a-a* marks the third person masculine singular of the perfect. A pair such as *katab* 'he wrote' and *ḍaṛaḇ* 'he hit' (both CA pausal forms) would require the *a-a* and an allomorph *a-a̤*. Another difficulty is trying to explain *ḍaṛaḇat* 'she hit', for the /b/ of *ḍaṛaḇ* 'he hit' is emphatic but the same /b/ of 'she hit' is non-emphatic.

Some Firthian linguists tried to impose a prosodic or suprasegmental approach on the Arabic phonological system (see, e.g., various papers in Mitchell 1975). This approach was apparently more satisfactory than those mentioned so far, since (as stressed earlier) emphasis never occurs as a monovalent phonemic feature. Rather, the minimum domain for emphasis in many Arabic dialects is the common syllabic type CV. Thus, emphasis can be seen as [+syllabic] or [-syllabic]. Critics of this analysis have pointed out

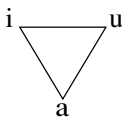
that one must also apply this frame of reference to voicing and velic closure as also being suprasegmental, since both appear to involve features that extend over more than one segment. However, when one given example occurs over the domain of a syllable, it is a distinctive feature of consonants but a redundant feature of vowels. This can be proven by noting the non-existence of nasalized or voiceless vowels outside nasal or voiceless environments. The corresponding situation for emphasis is quite different because it is not clear if consonants influence vowels or vice versa (as in the case of velic closure or voicing). Therefore, emphasis is much more suited for suprasegmental status than are the other two features mentioned above, since the latter extend over and are distinctive in only one segment.

Lehn (1963) reviewed much of the literature on the entire subject (including Arabic grammatical thought) and concluded that, at least for Cairo Arabic, the minimum domain of emphasis is the syllable, and the maximum domain is the utterance. Lehn suggested that emphasis not be treated as a distinctive system of the consonantal or vocalic system, but as a redundant feature of both. In his later works, Lehn underscores all emphatic syllables; thus, Cairene *ḍarab* 'he hit' but *ḍarabit* 'she hit'. This is testimony in favor of our notation *ḍarab* vs. *ḍarabit*.

As Ferguson (1956) has demonstrated, the /l/ which occurs only in the name of God in CA /ʔaḷlāh/ (but not after /i/ as in /bismillāh/ 'in the name of Allah') is a phoneme in CA. Some modern Arabic dialects, especially those spoken in the Gulf states, have many more examples of /l/.

13.5. Vowels

The Arabic vowel system is well known among linguists. It has the classical triangular system that preserves PS vocalism. This system is represented as:



For CA and MSA, the vowels may be short or long (geminated). Many modern Arabic dialects have, however, developed other vowels such as /ə/, /e/, /o/, etc., just as the other Semitic languages did centuries and millennia earlier through a general and natural process of parallel development.

Arabic vowel allophonics are much richer than its consonant allophonics chiefly because the former take on the coloring of the adjacent emphatic and emphatic-like consonants (including /r/), while the non-emphatic consonants

push the vowels to higher and less centralized qualities. Significantly, the pronunciation of MSA and any oral rendition of CA are all directly dependent on the nature of the speaker's colloquial dialect natively acquired. The influence of the native vernacular on MSA is most obvious in the matter of stress or consonants such as the *jiim* ([j]); however, detailed studies are sorely needed in this area, neglected by specialists for far too long.

The vowel allophonics have been accurately described on the basis of sound spectrographic analysis for MSA as used in Iraq (Al-Ani 1970). Very similar rules can be formulated for other parts of the Arab world. The rules may be stated as follows (where the lack of an underscore in the environment means that the segment may occur on either side of the motivating feature or segment):

- (1) /ī/ → [ī̄] / [+emphatic]_ (except /l/)
 → [ī̄] / _ { $\begin{smallmatrix} \text{ʕ} \\ \text{y} \end{smallmatrix} \}$ _
 → [ī] / (elsewhere)
- (2) /i/ → [ī] / _ [+emphatic]_
 → [i] / _ { $\begin{smallmatrix} \text{ʕ} \\ \text{y} \end{smallmatrix} \}$ _
 → [i] (elsewhere)
- (3) /ū/ → [ū̄] / _ [+emphatic]_ (except /l/)
 → [ū] (elsewhere)
- (4) /u/ → [ū] / _ [+emphatic]_
 → [u] (elsewhere)
- (5) /ā/ → [ā̄] / $\left[\begin{array}{c} \left\{ \begin{array}{c} \text{+emphatic} \\ \left\{ \begin{array}{c} \text{q} \\ \text{r} \end{array} \right\} \end{array} \right\} \\ - \end{array} \right] \right]$
 → [ā̄] / $\left[\begin{array}{c} \left\{ \begin{array}{c} \text{ʕ} \\ \text{y} \end{array} \right\} \\ - \end{array} \right]$
 → [ā̄] (elsewhere)
- (6) /a/ → [ə] / ___# (but not next to /q/, /ʕ/, /t/, and /y/)
 → [a] / $\left[\begin{array}{c} \left\{ \begin{array}{c} \text{+emphatic} \\ \left\{ \begin{array}{c} \text{q} \\ \text{r} \end{array} \right\} \end{array} \right\} \\ - \end{array} \right]$
 → [a] / $\left[\begin{array}{c} \left\{ \begin{array}{c} \text{ʕ} \\ \text{y} \end{array} \right\} \\ - \end{array} \right]$
 → [æ] (elsewhere)

In modern Arabic dialects, the short vowels are more susceptible to change than the long vowels. Thus, CA /i/ and /u/ in Damascene merge into /ə/ (exactly as occurred centuries earlier in Ge'ez—a good example of parallel development). Although /a/ is usually the most stable and conservative of the three short vowels, it too is now becoming subject to change or deletion in many dialects; e.g., Cairene /yā/ + /maḥammad/ → /ya mḥammad/ 'O Mohammed!'. Both Hebrew and Aramaic reduced the short vowels, particularly in unaccented environments. Cf. Hebrew *šəbīl* 'path' = CA *sabīl*, or Hebrew *kətabtém* 'you wrote (m. pl.)' = CA *katábtum*. CA has many doublets in its short vowel configuration such as /hubs/ ~ /hibs/ 'inalienable property the yield of which is devoted to pious purposes' or /laṣṣ/ ~ /liṣṣ/ ~ /luṣṣ/ 'thief', or /maḥjir/ ~ /miḥjar/ ~ /maḥjar/ 'eye socket' (triplets!).

There are two diphthongs in CA: /aw/ and /ay/, as in /θawr/ 'bull' and /bayt/ 'house', respectively. In most of the colloquial dialects, diphthongs have monophthongized into /ē/ and /ō/, respectively /i/ and /u/ in Moroccan dialects, as had also occurred in Akkadian millennia earlier—another well-attested example of parallel development). It should, however, be kept in mind that even in dialects which generally monophthongize the CA diphthongs, they are preserved intact in some words; e.g., Cairene *mawgūd* 'present (m.sg.); here'.

13.5.1. /a/-Raising

ʔimāla (lit., 'inclination') refers to /a/-raising, often due to the umlauting influence of /i/. A word such as CA *ʕibād* 'slaves' could have had a dialectal pronunciation *ʕibēd* or *ʕibīd*. *ʔimāla* has produced the very distinctive high-vowel pronunciations of /a/ in many Syro-Lebanese dialects, i.e., [bēb] or [bīb], equivalent to CA /bāb/ = [bæ:b] 'door' (as uttered, say, by a Cairene). Higher phonetic qualities can also be found, such as Maltese *bieb* [bɪ:p] 'door', cognate with the aforementioned word. Maltese also has *kelmiet* 'words', corresponding to CA *kalimāt* and *ktieb* 'book' = CA *kitāb*. Many dialects also have final *ʔimāla* observable in the feminine ending -a of MSA being pronounced -e or even -i. *ʔimāla* explains why French *Michelle* is written ميشال <myšʔl> in Arabic. It is interesting to note that *ʔimāla* is especially characteristic of women's speech (along with de-emphasis), both in their native colloquials as well as their renditions of CA and MSA. One may broadly compare *sodi* in Appalachian English for *soda*.

13.5.2. Delabialization

ʔiṣmām ‘delabialization’ is described as “the pronunciation of *u* with a trace of *i*” (Wehr 1974: 485). It describes the process in which /ū/ becomes /ĩ/ (perhaps through an intermediate stage of [ü]). This produces such alternations as *rūm* ~ *rīm* ‘Rome’ and some dialectal pronunciations of /rudda/ as [rüdda] ‘it was returned’ or /qūla/ for CA /qīla/ ‘it was said’ < a morphophonemic or underlying llquwilall. Incidentally, some Arabic dialects have [ü] (IPA [y])—e.g., Judeo-Moroccan and Mauretanian (Ḥassaniyya) Arabic, and Cairene is noted for [ü] (Harrell 1957: 5), IPA [y]. The front rounded vowel situation has yet to receive a full-length study in terms of comparative Arabic dialectology.

13.5.3. Stress

Stress is one of the most involved topics in the entire realm of Arabic phonology. Even for the Nigerian dialect of Arabic I researched firsthand in the mid-1970s, stress was the most intricate part of the variation I encountered (1982: xv). The medieval Arab grammarians never mentioned it, and modern-day stress of both MSA and CA is certainly often dependent on, and/or related to, the corresponding rules of the native colloquial dialect acquired. For instance, for the CA or MSA word <ktbt?>, segmentally /katabatā/ ‘both of them (f.) wrote’, some native speakers pronounce (1) /kátabatā/ (some Iraqis and Upper Egyptians), (2) /katabátā/ (Egyptians, especially Cairenes), (3) /katabatā/ (most Lebanese), and (4) /katábatā/ (most Jordanians and Palestinians). It is thus possible to stress any of the four syllables and still be absolutely correct. (This would be, as far as I know, a situation difficult to match in any other language.) This is but one of the many reasons for considering MSA to be an ill-defined system of language but all colloquials, on the other hand, to be well-defined systems. This distinction is at the heart of diglossia, really therefore multiglossia, in Arabic, a sociolinguistic situation best viewed as a continuum (see Kaye 1994).

There are, however, rules of syllabicity which can be described with a greater degree of accuracy. Long vowels are shortened in closed syllables (except /ā/ in some instances), which explains why one says /yákuṇ/ ‘let him be’ (jussive of /yakūnu/ ‘he will be’) instead of the expected (apocopated imperfect) */yakūṇ/. Also, only monoconsonants occur syllable-initially or -finally, rendering the borrowing of Latin *strata* ‘path’ as /ṣīrāt/.

The following rules for the assignment of lexical stress in MSA apply to the varieties of many Arabic users:

(1) When a word is made up of CV syllables, the first syllable receives the primary stress, e.g., /kátaba/ ‘he wrote’ (but Lebanese MSA *katábahu* ‘he wrote it’ vs. Egyptian MSA *katabáhu*).

(2) When a word contains only one long syllable, it receives the primary stress, e.g., /kátib/ ‘writer, clerk’.

(3) When a word contains two or more long syllables, the long syllable nearest to the end of the word receives the primary stress, e.g., /raʔisuhúnna/ ‘their chief (f.pl.)’.

13.6. Pausal forms in CA and MSA

The normal use of MSA and CA requires an understanding of a phenomenon traditionally known as pausal forms or pronunciations. When a pause occurs in speech (reflected in reading aloud as well), Arabic speakers drop final short vowels (case and mood markers). Furthermore, other endings are dropped or shortened. For example, Arabic marks indefiniteness by what is called *nunation* (named after the Arabic letter *nūn*): *-un* for nominative, *-in* for genitive, and *-an* for accusative (the only three cases). At the end of an utterance (i.e., sentence or breath group), a word such as /mudarrisun/ ‘a teacher’ → /mudarris/ and /mudarrisin/ → /mudarris/, but /mudarrisan/ → /mudarrisā/, whereas /mudarrisatun/ (nom.) or /mudarrisatin/ (gen.) or /mudarrisatan/ (acc.) → /mudarrisah/ ‘a teacher (f.sg.)’. Arabic words are not usually cited with nunation (called in Arabic *tanwīn*), e.g., *kitābun* ‘a book’ because isolated words (e.g., *kitāb*) are associated with pause. Native vernacular influence is less obvious in word-isolated speech (single citation forms) than it is in long, connected discourse. It is important to note that, in principle, vernacular forms are based on CA pausal forms. It is important to note that no one currently speaks MSA utilizing all the CA pausal rules.

13.7. Morphophonemic changes

There are far too many morphophonemic changes to include in our discussion here. A few of the most common occurring in CA are as follows (data are cited in non-pausal forms):

- (7) *awa* → *ā*
 **qawama* → *qāma* ‘he stood up’
- (8) $C^1aC^2aC^2a \rightarrow C^1aC^2C^2a$
 **radada* → *radda* ‘he returned’

- (9)
$$\mathcal{P} \left\{ \begin{matrix} a \\ i \\ u \end{matrix} \right\} \mathcal{P} \rightarrow \mathcal{P} \left\{ \begin{matrix} \bar{a} \\ \bar{i} \\ \bar{u} \end{matrix} \right\}$$

**ʔaʔlāmūn* → *ʔālāmūn* ‘pains’
- (10) *uw* → *ū*
**suwdun* → *sūdun* ‘black (m.pl.)’
- (11) *ūy* → *ī*
**buyḍun* → *bīḍun* ‘white (m.pl.)’
**mudarrisūya* → *mudarrisīya* ‘my teachers (m., all cases)’
- (12) *yw* → *yy*
**ʔaywāmūn* → *ʔayyāmūn* ‘days’
- (13) Haplology: **tataqāṭalūna* → *taqāṭalūna* ‘you are fighting each other’
- (14) Dissimilation: **madīnīyun* → *madanīyun* ‘urban’
- (15) *āw* → *āʔ*
**qāwīlun* → *qāʔīlun* ‘speaker’

Arabic’s basic triconsonantal root structure lends itself well to the principles and techniques of what is called “underlying” phonological analysis. Thus, rule (7) states that the root *qwm* plus the discontinuous vocalic pattern *a-a-a* produces *qawama* (yielding eventually *qāma* ‘he got up’). One can see the *w* or second radical of the root in the imperfect corresponding to this aforementioned form, viz., *yaquwmu*, in which *ya-* marks the third person and *-u* the indicative mode. Rule (10) operates on this underlying representation, yielding the surface manifestation *yaqūmu*. The root structure allows for the almost algebraic formulae which can describe various aspects of Arabic morphophonology. For instance, as one looks at the root KTB ‘write’, one can readily describe MSA *kataba* ‘he wrote’, *kātibun* ‘clerk’, *kitābun* ‘book’, *kutubun* ‘books’, *maktabun* ‘office; desk’, *maktabatun* ‘library’, *mukātabatun* ‘correspondence’, and so on. Cf. *qāla* ‘he said’, *qāʔīlun* ‘speaker’, *maqālun* ~ *maqālatun* ‘article, essay’, *muqāwalatun* ‘conversation’, *qawl*, pl. *ʔaqwāl* ‘saying’, etc.

13.8. Some phonological rules in Cairo Arabic

Finally, no discussion of Arabic phonology would be complete without a brief look at some major phonological rules in a spoken colloquial dialect.

(It should not be forgotten that no one speaks MSA or CA as a mother tongue.) Cairo Arabic is chosen for two reasons: (1) It is relatively well researched (see Mitchell 1956, 1962, 1990, 1993; Abdel-Massih 1975; and Salib 1981); and (2) many Arabs are at least somewhat familiar with this dialect due to its political and cultural importance (particularly from the movies since Cairo is the film capital of the Arab world, and attributable to the overwhelming influence of Egyptian radio and television throughout the Middle East). Due to limitations of space, the tremendous variety of Arabic phonologies is not the main focus of this chapter. On the latter topic the reader is referred to chapters 14, 15, and 16.

13.8.1. *Anaptyxis or epenthesis*

Cairene does not allow a sequence of three consonants (including geminates, keeping in mind that any consonant can be geminated). When -CC#C- occurs (i.e., a consonant cluster at the end of a word followed by another consonant), the anaptyctic vowel /i/ is usually inserted and is pronounced as an ultrashort [ɪ] (= [ɪ]). For example, /dars/ 'lesson, class' + /ʕarabi/ 'Arabic' becomes /darsi ʕarabi/ 'Arabic class'. Some scholars have transcribed this vowel as /i/ with a breve mark. Before some of the pronominal suffixes, the anaptyctic vowel can be /a/ or /u/, usually determined by vowel harmony, viz., /u/ before /-hum/ or /-kum/ (3pl. and 2pl. respectively) and /a/ before /-ha/ (3f.sg.). Thus, /gamb/ 'beside' becomes /gambuhum/ 'beside them', /gambukum/ 'beside you (pl.)', /gambaha/ 'beside her'; but note /gambina/ 'beside us', which is not vowel harmonic. The state of affairs described above is in sharp contrast with the situation of Moroccan Arabic, in which many consonants can occur adjacent to one another in clusters (see chapter 14). The facts of Moroccan Arabic phonology are to be explained by a Berber substratum, in which many consonants can occur next to one another without any audible vowel (e.g., Zuara Berber).

13.8.2. *Consonant assimilation*

Consonant assimilation in Cairene occurs quite often in rapid speech (but it may occur in careful speech as well). Examples include /dt/ → /tt/, /td/ → /dd/, and /ʔil-C/ → /ʔiC-C/. Anticipatory assimilation usually occurs at word boundaries, especially for change in voicing according to the consonant which follows: /z/ → /s/, /ɣ/ → /x/, and /ʕ/ → /h/. Other common assimilations include /hh/ → /ħħ/, /ʕh/ → /ħh/, /sʃ/ → /ʃʃ/, and /ʕs/ → /ss/. An assimilation such as /simiʕt/ → /simiht/ 'I heard' clearly demonstrates that /ʕ/ and

/ħ/ are a structurally similar pair. One may note /simiħ/ 'he heard' + /-ħa/ 'her' → /simiħħa/ 'he heard her' with a progressive assimilation of /ħ/ to /ħħ/ (which operates after the rule which shifts /s/ → /ħ/).

13.8.3. Vowel shortening

One of the most characteristic phenomena of Cairene is vowel shortening, which occurs under predictable conditions:

(1) A long vowel in a final syllable shortens if liaison with a following word causes it to be followed by -CC, e.g., /šāfū hināk/ → /šāfu hnāk/ 'they saw him there'.

(2) A long vowel shortens if another long vowel follows it, e.g., /kitāb/ 'book' + /-ēn/ (dual) → /kitabēn/ 'two books', or /kātib/ 'clerk (m.)' but /katba/ 'clerk (f.)' < /kātiba/ (some other dialects keep the long vowel in the derivative formations).

(3) A long vowel becomes short when it is followed by two consonants, e.g. /bēt/ 'house' + /-na/ 'our' → /bitna/ 'our house' (for */betna/; note /e/ → /i/ via vocalic neutralization).

(4) A long vowel shortens if suffixation puts it before the penult, e.g. /šāfit/ 'she saw' + /-u/ 'him' → /šāfitu/ 'she saw him'.

13.9. Further reading

The reader is referred to five excellent works on Egyptian Colloquial Arabic (the last two deal with many other Arabic dialects as well): Harrell 1957, Mitchell 1956, 1962, 1990, 1993. The following two pedagogical works are also quite useful: Abdel-Massih 1975, Salib 1981. The following have been quoted in Table 13-1 and are most valuable for the listing of Semitic cognates: Köhler and Baumgartner 1958, Leslau 1987.

References

- Abdel-Massih, Ernest T. 1975. *An Introduction to Egyptian Arabic*. Ann Arbor: University of Michigan.
- Al-Ani, Salman H. 1970. *Arabic Phonology: An Acoustical and Physiological Investigation*. The Hague: Mouton.
- Ferguson, Charles A. 1956. "The Emphatic *l* in Arabic." *Language* 32: 446–52.
- Gairdner, W. H. T. 1925. *The Phonetics of Arabic*. London: Oxford University Press.

- Harrell, Richard S. 1957. *The Phonology of Colloquial Egyptian Arabic*. New York: American Council of Learned Societies.
- Hetzron, Robert. 1972. *Ethiopian Semitic: Studies in Classification* (Journal of Semitic Studies Monograph 2). Manchester: Manchester Univ. Press.
- Jakobson, Roman. 1957. "Mufaxxama: The 'Emphatic' Phonemes in Arabic." In *Studies Presented to Joshua Whatmough on His Sixtieth Birthday*, ed. Ernst Pulgram, pp. 105–15. The Hague: Mouton.
- Kaye, Alan S. 1972. "Arabic /žiim/: A Synchronic and Diachronic Study." *Linguistics* 79: 31–72.
- . 1982. *A Dictionary of Nigerian Arabic* (Bibliotheca Afroasiatica 1). Malibu: Undena.
- . 1994. "Formal vs. Informal Arabic: Diglossia, Triglossia, Tetraglossia, etc., Polyglossia—Multiglossia Viewed as a Continuum." *Zeitschrift für arabische Linguistik* 27: 47–66.
- Köhler, Ludwig, and Walter Baumgartner. 1958. *Lexicon in Veteris Testamenti Libros*. Leiden: Brill.
- Lehn, Walter. 1963. "Emphasis in Cairo Arabic." *Language* 39: 29–39.
- Leslau, Wolf. 1987. *Comparative Dictionary of Ge'ez*. Wiesbaden: Harrassowitz.
- . 1990. *Arabic Loanwords in Ethiopian Semitic*. Wiesbaden: Harrassowitz.
- Mitchell, T. F. 1956. *An Introduction to Egyptian Colloquial Arabic*. London: Oxford University Press.
- . 1962. *Colloquial Arabic: The Living Language of Egypt*. London: English Universities Press (Teach Yourself).
- . 1975. *Principles of Firthian Linguistics*. London: Longman.
- . 1990. *Pronouncing Arabic I*. Oxford: Oxford University Press. [This work and Mitchell 1993 deal with many Arabic dialects, Modern Standard Arabic, and Classical Arabic.]
- . 1993. *Pronouncing Arabic II*. Oxford: Oxford University Press.
- Nelson, Kristina. 1985. *The Art of Reciting the Qur'an*. Austin: University of Texas Press.
- Owens, Jonathan. 1993. *A Grammar of Nigerian Arabic* (Semitica Viva 10). Wiesbaden: Harrassowitz.
- Salib, Maurice. 1981. *Spoken Arabic of Cairo*. Cairo: American University of Cairo Press.
- Wehr, Hans. 1974. *A Dictionary of Modern Written Arabic*, ed. J Milton Cowan. Wiesbaden: Harrassowitz.

Moroccan Arabic Phonology

Jeffrey Heath

University of Michigan

14.1. Introduction

Moroccan Colloquial Arabic (MCA) is one of many dialects (actually, dialect groups) that resulted from Arab military and cultural expansion shortly after the time of Muhammad. Because Arabic was overlain on various foreign languages in the course of this expansion (Persian in Iraq, Aramaic in the Levant, Coptic in Egypt), the respective colloquial vernaculars are quite different—from each other, from the modern colloquials of the Arabian peninsula, and from Classical Arabic (CA) and the modern literary Arabic based on it. Though all of the colloquials are referred to as “Arabic” and are regarded by most Arabs as low-status street-level varieties of a single literary language, they are mutually unintelligible and therefore (from the point of view of linguistic description) distinct languages.

MCA is really a collection of dialects, sharing some important features (in contrast to, say, Egyptian) but with some surprisingly important internal differences in all aspects of lexicon and grammar. The variation is, for the most part, not home grown, but rather a result of the complex settlement history of Morocco. The major components are these: (a) core dialects reflecting early Arabization of Berbers (from around 700 A.D.), (b) an invasion of “Beni Hilal” from Arabia (mid-11th century), and (c) an influx of Muslims and Jews from Spain (around 1492).

When the Arabs overran Berber-speaking Northwest Africa in the early 8th century, they established military camps, then towns, and progressively Arabized many of the Berbers. This slow Arabization process is still continuing, as the remaining Berber populations (especially those who relocate to urban centers) become bilingual and shift toward Arabic colloquials. In Morocco, unlike most of the rest of the Maghreb, the Berber languages are still doing well in the mountainous areas.

Although MCA is the unquestionably dominant vernacular in Morocco, it has many traces of Berber influence. Since Arabized Berbers must have constituted a large percentage of speakers in the formative first centuries of

MCA, it is no wonder that its pronunciation in particular is quite different from that of eastern Arabic vernaculars such as Egyptian colloquial. Two notable characteristics of MCA are the hardening of interdental fricatives (**ð*, **ḏ*, and **θ*) to stops (**d*, **ḏ*, and **t*), and the reduction (to shwa) or complete loss of original CA short vowels. The latter development has had far-reaching consequences for the entire phonology, prosody, and morphology.

The so-called “Beni Hilal,” whose precise tribal and geographical origin in the Arabian Peninsula are not entirely clear, were unleashed onto north-west Africa by an angry Egyptian sultan. They were speakers of a “nomadic”-type dialect (the sedentary/nomadic division of dialects is an old theme in Arabic dialectology). Many of them passed through the Maghreb and kept going—into the Sahara, and beyond to Mauritania and northern Mali. Their descendants are the Moors of these two countries, and of Western Sahara (which Morocco is now attempting to annex), and their Arabic dialect (called Hassaniya) differs sharply from MCA in phonological segments, phonology (e.g. *i-žu* ‘they are coming’ vs. MCA *y-ži-w*), morphology (e.g. preservation of feminine plurals), and lexicon. However, some of the Beni Hilal remained in the Maghreb, north of the Sahara. They must have been partly responsible for the strongly “nomadic” features of Tunisian Arabic and of some Algerian varieties. In the settled regions of Morocco, the Beni Hilal formed smaller pockets, exacerbating the sedentary/nomadic split, but many “nomadic” features have since been leveled out. Phonological characteristics of the remaining “nomadic” dialects (best represented in the southern oases and the Saharan fringe) include retention of CA interdental fricatives **ð*, **ḏ*, and **θ* (which become stops elsewhere in MCA), old **q* reflected as *g*, and retention of (originally long) CA vowels in the final stem-syllable of broken plurals and diminutives: *brrad* ‘tea kettle’ has pl. *brarid* and diminutive sg. *breyrid* in the far south, versus *brar(ə)d* and *brir(ə)d* elsewhere in Morocco.

The expulsion of Jews from Spain in 1492 and the various expulsions or outmigrations of Andalusian Arabs around the same time (the last Arab kingdom, Granada, fell in 1492) had important consequences for Arabic dialects in the Maghreb, and especially in Morocco. The Andalusians, more sophisticated economically and culturally than the Moroccans, formed their own neighborhoods and networks in Moroccan cities, and in some (such as Fes) there are still distinctive elite urban dialects with Andalusian features, though dialect leveling is taking an increasing toll on them. The Jews, who came from the Christian as well as the Muslim parts of Spain (they still have

family names like Toledano), spoke Judeo-Spanish. Though there have been Jewish communities in Morocco since the beginning of recorded history, the more sophisticated Spanish Jews dominated the major Jewish communities, especially in Rabat, Meknes, Fes, and Sefrou. In the far north (Tangier, Tetuan) they continued to speak Judeo-Spanish into the early 20th century. Elsewhere they shifted to MCA, specifically the then-current urban dialects of Jews and Muslims, with some Andalusian Arabic influence. When the Jews were then ghettoized in walled-off urban quarters, they became cut off from urban/rural blending among Muslims and still preserve many archaic features that even the old Muslim urban neighborhoods have since given up. Moroccan Judeo-Arabic (MJA) therefore constitutes a second dialect network, occupying essentially the same geographical space as the more visible Muslim dialects. (Because of recent migrations to Israel and consolidation of remaining Moroccan Jews chiefly in Casablanca, the old Judeo-Arabic dialects are in the process of dying out as vernaculars.)

Phonological features typical of at least the major urban MJA dialects include (a) retention of CA **q* as glottalized *q'* or glottal stop *ʔ* rather than as *g*; (b) loss or unrounding of short **ū*, with no transfer of rounding to neighboring back consonants; (c) loss of the alveolar/palato-alveolar distinction among sibilants (**s* merges with **š*, **z* with **ž*); and (d) spreading of pharyngealization (“emphasis”) through lexical stems. However, MJA dialects occur almost throughout Morocco, and those in the far south and far east diverge from these patterns in some respects. The MJA dialect of Tafilalt, a semi-arid region in the south, is notable for its extreme reduction or loss of CA short vowels.

In what follows, unless otherwise specified the “MCA” that we are describing is the speech of Moroccans born after about 1950, living in or near the central urban belt (from Fes to Rabat) but socially mobile and with no especially prominent indicators of local birthplace. The present article, based on Heath (1987), does not pretend to be dialectologically comprehensive. A sketch of the dialectal networks is in Heath (1991); a fuller treatment is in preparation.

14.2. Consonants

The basic consonant system of MCA is as shown in Table 5-1. Found dialectally are *ʔ* and the fricative trio *ð*, *ð̤*, and *θ*. Foreign loanwords, especially from French, have introduced *j* (affricate), *v*, and *p* into MCA.

Table 14-1. Primary Moroccan Colloquial Arabic Consonants

	(Bi-)labial	(Denti-)alveolar	Pharyngealized (denti-)alveolar	Palato-alveolar	Velar	Uvular	Pharyngeal	Laryngeal (aspiration)
Voiceless stops/ affricates		t	ɸ	ʈ	k	q		
Voiced stops/ affricates	b	d	ɖ		g			
Voiceless fricatives/sibilants	f	s	ʃ	ʂ		x	ħ	h
Voiced fricatives/sibilants		z	ʒ	ʐ		ɣ	ʕ	
Nasals	m	n						
Laterals		l	(ɭ)					
Rhotics		r	ɾ					
Semivowels	w			y				

All consonants may occur geminated, and gemination is a productive part of derivational ablaut: *y-bki* ‘he weeps’, causative *y-bkki-ha* ‘he makes her (-ha) weep’. Definite *l-* (cf. CA *al-*) before a stem-initial coronal (alveolar or palato-alveolar) loses its own identity to form a geminate cluster: *l-brd* ‘the cold’, but *t-ṭrab* ‘the ground’ and *ʒ-ʒbəl* ‘the mountain’. The preposition *l-* ‘to’ (CA *li-* or *ʔilaa*) does not assimilate in this fashion: *l-ṭanʒa* ‘to Tangiers’. Note that *ʒ* is treated like any other coronal in the definite assimilation rule, whereas its ancestor (CA *j*) did not assimilate (*al-ʔabal* ‘the mountain’). In far northern MCA dialects, a geminated *ʒ* is realized as [j:], so /ʒ-ʒbəl/ is [j:ːbəl].

14.3. Full vowels

The vowel system differs substantially from one sub-dialect to another, but the stable core of the vocalic system is the set of “full” vowels *i*, *a*, *u*. In general, these are historical reflexes of long vowels and diphthongs in CA, but since there is no longer a set of parallel short vowels it is not appropriate to

consider length as part of the MCA system.

What happened to the old CA short vowels? Many have simply disappeared. Word-final short vowels were unstable even in CA, where they were omitted before e.g. a pause. In the development from CA to MCA, not only did the word-final short vowels complete their disappearing act; many other short vowels in other positions were deleted due to syncope, like the first vowel of CA *jabal* → MCA *žbəl*. Short-vowel prefixes like CA *ta-* ‘you or she (imperfective)’ are now typically reduced to unvocalized consonants, which may undergo assimilation to a stem-initial consonant: CA *ta-ḍrib(-u)* ‘you (m. sg.)/she hits’ → MCA *ḍ-ḍrb*.

14.4. Vestiges of short vowels

There are, however, some traces of the old CA short vowels. In the mainstream MCA dialect we are focusing on, old short **i* and **a* are reflected in certain positions as ə, while short **u* is reflected variously as shwa or as a short rounded MCA vowel that may be written *ũ*. The position most favorable to these non-zero reflexes of old short vowels is before the final consonant (or two-consonant cluster) of a word not containing a full vowel, i.e., positions like C_CC and CC_C, where “C” represents any consonant. MCA preserves no word-final short vowels, nor are there any clear cases of preservation word-initially.

Depending on sub-dialectal details, MCA short *ũ* and perhaps even ə can be re-analysed as something other than ordinary short vowels. In the mainstream dialect, *ũ* has a distribution like that of shwa, except that *ũ* always seems to be in the vicinity of a velar or uvular consonant, hence *skũt* ‘shut up!’, *xũbz* ‘bread’, and *gũl-t* ‘I said’. This association of *ũ* with particular consonants is an innovation in MCA. There are two other interesting facts about MCA *ũ*. First, paradigmatic alternations between rounded and unrounded short vowels have generally been leveled out; the CA verb ‘to shut up’ had perfective *sakat-* vs. imperfective *-skut-*, which would normally have produced an MCA opposition **skət/skũt*, but in fact *skũt* has generalized to the perfective: *skũt* ‘he shut up’, *skũt-t* ‘I shut up’. Second, in positions where MCA shwa undergoes syncope, as in *šəmm* ‘smell! (m. sg.)’ but *šmm-i* ‘smell! (f. sg.)’, short *ũ* typically leaves labialization behind as a trace: *dũqq* ‘knock! (m. sg.)’ but *dʷqq-i* or *dqqʷ-i* ‘knock! (f. sg.)’. What these facts seem to point to is the increasingly close association of rounding with the back (velar or uvular) consonant rather than with the (at best faint) short vowel. In other

words, a form like [skūt] might well have been re-analysed by Moroccans as /sk^wət/, with a phonemic *k^w* that accounts for the rounding. Under this (re-)analysis, we can think of the vowel as an underlying shwa.

If there is only one short vowel, ə, versus three full vowels, *i a u*, it is obvious that a length opposition in the usual sense no longer exists—especially in view of the very limited set of positions where shwa can occur. At this point, one is tempted to ask whether even this restricted shwa can be “analysed out” of the system. If, for example, it occurs only in the positions C_CC and CC_C, we could consider the possibility that shwa is inserted by rule rather than being part of a lexical representation. This is probably going too far, since the very fact that CəCC and CCəC are distinct stem shapes shows that no simple shwa-insertion rule is possible. However, it may well be that this is the long-range developmental trend. In mainstream MCA, CəCC and CCəC no longer contrast when the third C is a sonorant (or when the second C is a rhotic). In some dialects of the Tafilalt area in the south, the distinction is additionally lost when the second C is any sonorant, and in the same dialects we are starting to get sporadic “crossovers” even with obstruents (*xbəz* instead of *xǔbz* or *xəbz* ‘bread’). In such dialects, ə has virtually disappeared as a structurally significant phoneme, leaving *i a u* as the only vowels in the system.

14.5. Pharyngealization (“emphasis”)

Kaye’s article on Arabic (this volume) summarizes the debate about the phonological representation of “emphatic” consonants (or vowels, or syllables, or words). That the various dialects themselves tend toward different interpretations of this problem can be shown by a comparison between mainstream MCA and the Saharan dialects of the extreme south (and especially of Western Sahara, Mauritania, and Mali).

In the Saharan dialects, the “spreading” of pharyngealization is very narrow. Typically a pharyngealized C affects the neighboring vowel and nothing else: in *y-ǣsgāl* ‘he cleans’, the first vowel is backed under the influence of the pharyngealized *ṣ*, but the second vowel is unaffected, hence [yǣsgǣl]. In mainstream MCA, on the other hand, spreading is more substantial, especially to the “right”: *tlb-at-u* [tlb-at-u] ‘she requested it’ has a clearly backed allophone of *a*. In addition, in Saharan dialects the phonetic effect of pharyngealization on an adjoining high vowel *i u* is an evanescent velarization, marked strongly in the actual transition between the vowel and the pharyn-

gealized C and then fading out, whereas MCA *i u* are pronounced as steady-state mid vowels [e o] next to pharyngealized Cs.

Related to this is the fact that mainstream MCA tends to avoid combinations, in the same stem, of a basic pharyngealized alveolar *t d ʂ z ɾ* with an unpharyngealized alveolar *t d s z r*—alveolars in the same stem are either all pharyngealized or all plain. Thus CA *ʂadr* ‘chest (of body)’ becomes *ʂɗr* in some Moroccan dialects, *sdr* in others, but mixtures like **ʂdr* are rare (except in artificially re-classicized pronunciations). By contrast (leaving aside *r* for the moment), Saharan Arabic seems to prefer having just one pharyngealized alveolar in a stem, and there are cases where a consonant has lost its original pharyngealization in proximity to a neighboring pharyngealized or uvular C. Thus original CA *ʂayīr* ‘small’, which becomes mainstream MCA *ʂyīr* with pharyngealization extending to the rhotic, produces *syīr* in Saharan Arabic, the uvular fricative having prevented secondary pharyngealization of the rhotic and having actually depharyngealized the sibilant.

Because MCA spreads pharyngealization at least throughout stems to a much greater extent than Saharan (or most other Arabic vernaculars), the argument that “emphasis” is a lexical feature of stems rather than of individual consonants has more weight in MCA than elsewhere. It is even possible to experiment with radical re-interpretations—e.g. to take vocalic distinctions as primary, leading perhaps to an analysis in terms of two “vowel harmony” sets of full vowels, *i æ u* vs. *e a o*. There is no strong evidence that MCA has embraced this re-analysis, but it may be lurking in the shadows.

14.6. *r* versus *ɾ*

In CA, there appears to have been just one rhotic phoneme **ɾ*, though it probably already had some allophonic variation depending on vocalic and consonantal environment. In the Maghrebi and Saharan (Hassaniya) dialects, a distinction between *r* and *ɾ* has now become phonemic. We can easily see how this opposition arose by considering *kbr* ‘bigger’ (comparative of *kbīr* ‘big’) and *kbr* ‘(to) grow, become big’. The immediate (pre-MCA) etyma probably had the forms **kbār* ‘bigger’ and **kbīr* ‘(to) grow’, respectively. The *a*-vocalism favored a “dark” allophone of **ɾ*, while *i*-vocalism favored a “light” allophone. When the old short vowels merged (and were then swallowed up by the rhotic, which has become syllabic in these forms), the originally allophonic distinction became phonemic.

Although MCA has a few examples like *kbr/kbr*, apparently involving forms of the “same” root, such alternations are rare and it is attractive to interpret the forms as lexically distinct. Since there are only a few (about six) remaining morphological comparatives like *kbr*, they are probably best regarded as lexicalized (some, like *ḥsən* ‘better’ from *mzyan* ‘good’, are suppletive). In fact, the tendency has been to level out *r/ṛ* alternations, generalizing one or the other within the paradigms and derivatives of each stem. This leveling has gone farther in Morocco than elsewhere in northwest African Arabic.

In Saharan (and Tunisian) Arabic, despite some paradigm leveling there are still many telltale alternations. In particular, active participles (classical *CaaCīC) typically depharyngealize *r* to *r* in C₂ or C₃ position if no other pharyngealized C is present: *mṛāg* ‘he went away’, participle *maarag* (**maarig*). In mainstream MCA, by contrast, verbs with *r* generalize it to participles: *ṣṛəb* ‘he drank’, *ṣaṛ(ə)b* (**ṣaarib*) ‘having drunk’. Contrast the noun *ṣaṛ(ə)b* ‘lip’, which shows the regular reflex of this same **ṣaarib* (here a specialized nominal usage, meaning ‘moustache’ in CA) when not subject to analogical influence from the verb ‘to drink’. (See also § 14.8.)

14.7. Derivational ablaut

Like other forms of Arabic, MCA has a system of stem-derivation whose most obvious effect is changes in vowels: *ktab* ‘book’, pl. *ktuba*, diminutive sg. *k^wtiyyəb*, verb *ktəb* ‘(to) write’, active participle *katəb* ‘having written’ (or agentive: ‘writer’), passive participle *m-ktub* ‘written’, etc. The early Arab grammarians, and many present-day linguists, interpret such data to mean that lexical roots are consonantal skeleta like *k-t-b*, which are interspliced with grammatically significant vowel patterns to produce actual stem shapes. In the currently most popular model, both the consonantal skeleton and the grammatical vowel sequence are linked (separately) to a third component—a “template” of consonantal and vocalic positions, such as CVCC or *m*-CCVC. In this model, a form like *katəb* ‘having written’ would involve a template approximately of the shape CVCC, a root skeleton *k-t-b*, and a vocalic pattern *a* (or *a-ə* if the shwa is recognized as part of the rule). Such models require a complex set of rules for linking the root consonants and grammatical vowels to C and V positions of the template.

This model radically distinguishes the morphology of Arabic from the transparently layered morphology of other languages, where an “inner”

stem X is the basis for a suffixally derived stem [X]-S₁, which in turn is the source for other prefixally and/or suffixally derived stems P₁-[[X]-S₁], then [P₁-[[X]-S₁]-S₂, etc. In an oversized English derivative like *antidisestablishmentarianism*, it is generally possible to identify a core stem (*establish*) and an ordered (bracketable) set of derivational affixes. A root/pattern analysis of MCA or other forms of Arabic would have the power to shortcut this layering, whereby each surface stem form could be directly referred to an abstract consonantal root, not to an intervening prior stem shape.

In practice, phonologists working on Arabic with the root/pattern model have had to incorporate some prior stem-shape information in their rules. In particular, the shape of singular nouns is obviously relevant to the shapes of diminutives and plurals, and at least the CV-template shapes of the singular have been allowed to play a role in generating these derivatives. However, all of the derivational patterns with well-defined templatic shapes appear to be asymmetrical, in the sense that one stem is clearly derived from another—the participles are derived from the corresponding verbs, just as the diminutive or the plural is from the singular noun. It is therefore possible in principle to trace each derived form back, directly or via intervening derivations, to a single source stem—usually a singular noun or adjective, or a simple verb.

These underived nouns, adjectives, and verbs tend to have unpredictable (and therefore, arguably, lexical) templatic shapes and vowel qualities, in contrast to the many derivatives with their rigidly prescribed vowels and templatic shapes. It seems reasonable, therefore, to analyse the underived stems as ordinary strings of consonants and vowels, just as in other languages, without the dissection into consonantal “roots,” vocalic “patterns,” and CV “templates.” In MCA, we would represent simple stems as *xūbz* ‘bread’ (or /x^həbz/, cf. § 14.4), *ktab* ‘book’, and *ktəb* ‘(to) write’. Although suggestively similar in sound and meaning, the last two are not relatable by any productive derivational rule (CCaC as a nominal shape is uncommon and has no specific grammatical value) and so are both treated as underived; in some ancient form of pre-Arabic they may have been derivationally related. (Jewish MCA dialects typically replace *ktab*, which has scriptural associations, with *mšhaf*, leaving the verb *ktəb* unaffected.)

Many stems, however, are clearly derived by productive stem-formation processes that we refer to collectively as “derivational ablaut.” In MCA, each derived stem is produced by the interaction between an input stem (either underived, or due to a prior derivation) and a template. Since the vowel

qualities and the CV shapes are tightly fused, we may analyze the derivational templates as consisting of pre-specified vowels and blank consonantal positions (represented by “C”). For example, the active and passive participles (for verbs like *ktəb* with exactly three segments, excluding “short vowels” like shwa) are *CaCC* and *m-CCuC*, respectively, and they combine with *ktəb* ‘(to) write’ to produce *kat(ə)b* ‘having written’ and *m-ktub* ‘written’.

Derivational ablaut, like affixal derivation in other languages, can be layered, though the layering is less transparent in MCA. From *dxül* ‘(to) enter’ we first create a causative/factitive *dxxəl* ‘(to) insert, put in’. This in turn becomes the input to affixal derivatives like participle (active or passive) *m-dxxəl*, and to further ablaut derivation, e.g., verbal noun *t-dxil* (or *t-dxal*) ‘insertion’ (compare verbal noun *dxul* ‘entrance’ from the underived verb).

Within what we may call the “stem/ablaut” model, in contrast to the root/pattern model, we (a) simplify the phonological representation of underived stems, and (b) recognize layering (derivational sequencing). In both respects, the stem/ablaut model treats MCA as not very different from other (affixing) languages. But like the root/pattern model, the stem/ablaut model does require some fairly intricate phonological subrules.

Consider what happens when an input stem and a derivational template constitute a mismatch—the input stem has too few consonants, for example. While *ktəb* is a good match for *CaCC* and *m-CCuC*, stems like *-akül* ‘eat’, *kun* ‘(to) be’, and *qra* ‘(to) read, study’ are not. The active participles are in fact *wakəl*, *kayən*, and *qari* (feminine *qary-a*). The type *wakəl* is found only with two verbs whose imperfective stems begin with *a* (in both cases a CA stem-initial *ʔ has been lost); these two stems fill the first consonantal position of all their ablaut derivatives with a *w* that may be regarded as secondary. Verbs like *kun* with imperfective *CuC*, *CiC*, or *CaC* have active participle *CayəC*. Since we get *y* in the participle regardless of the input vowel quality (*a*, *i*, or *u*), it appears that this semivowel is inserted by rule after the initial and final input consonants are transferred to the corresponding positions of the derivative. Verbs like *qra* with imperfective shape *CCa*, *CCi*, or *CCu* all have the shape *CaCi*, which is shown to be really /CaCy/ by feminine *CaCy-a*. Since other ablaut derivations, including the causative, merge verb-final vowels as *i*, it is possible that this participle reflects a similar neutralization, with the *i* then being transferred to the final C position of *CaCC* and hence being desyllabified to the semivowel *y*. These examples suggest that input vowels, not just consonants, play a role in determining the output, and this

is seen more transparently in cases where input stems with *u* and *i* have distinct outputs with *w* and *y*, respectively: *šuf* ‘(to) see’, reciprocal *t-šawf*-, versus *biŷ* ‘(to) sell’, reciprocal *t-bayŷ*-.

As these few examples suggest, even when one has chosen a basic morphological model (such as the stem/ablaut model here advocated), there are a great many phonological subtleties to work out.

14.8. Loanwords

We observed in § 14.5 that MCA has added *r/ɾ* to the set of well-established plain/pharyngealized consonantal oppositions inherited from CA. We pointed out that this opposition originated in allophonic variants due chiefly to vocalic environment (**i* favoring plain *r*; back vowels favoring *ɾ*), and that the reduction or disappearance of old short vowels led to the phonemicization of the *r/ɾ* opposition, as in *kbr* ‘bigger’ vs. verb *kbr* ‘(to) grow’.

We pointed out that most such alternations (involving the same stem) have by now been leveled out in MCA, so that the trend is for each stem (and its associated “family” of inflected and derived forms) to generalize either *r* or *ɾ*. There is, however, one productive derivational ablaut that seems to have resisted this leveling.

The most common plural corresponding to singular nouns of the shape *CaC* is *CɿC-an* (*wad* ‘river’, pl. *wid-an*). There are a number of native MCA nouns with a rhotic as final *C*, and here we consistently get *ɾ* in the singular versus plain *r* in the plural (the latter undoubtedly due historically to the preceding *i*-vowel): *faɾ* ‘mouse’, pl. *fir-an*. Only in the case of *ɣaɾ* ‘(mouse) hole, burrow’ do we find sub-dialectal variation between pl. *ɣir-an* and *ɣiɾ-an* (about evenly distributed), and here the uvular fricative has apparently been the differential factor. One would conclude from this inspection of inherited forms that the *r/ɾ* opposition is quite stable in this singular/plural pattern.

However, loanword data show conclusively that this is not so. There are two twentieth-century loanwords, *kaɾ* ‘intercity bus’ and *baɾ* ‘bar, tavern’, which match the pattern of *faɾ* ‘mouse’ and lack uvulars, so we would expect plurals **kir-an* and **bir-an* with plain *r*. Instead, we get *kɿɾ-an* and *bɿɾ-an* with pharyngealized *ɾ*. What this reveals is an otherwise hidden pressure on the *faɾ/fir-an* pattern to extend *ɾ* from singular to plural—a pressure that has heretofore been unable to actually reshape the inherited plurals, but

which has succeeded in affecting new (singular) loanwords with no previously established MCA plural.

This example is typical of the subtlety of “loanword phonology” in a language like MCA with a complex and sometimes murky ablaut system. Not only must each loanword be assimilated to MCA morphology and phonology in its “foot-in-the-door” form; the borrowing is then fed into the productive ablaut processes—which in turn may require deceptively complex analytical “decisions” by MCA speakers (Heath 1989).

While the student of language contact phenomena must deal with many tricky cases like this one, involving productive derivations and fairly ordinary borrowed stems, there is also an occasional poetic jewel that comes along. Most young Moroccans are familiar with an MCA noun *xwadr-i* meaning ‘brother’ or ‘pal, buddy’, but they have no idea of its origin and it is up to the linguist to unravel it. The ingredients are (a) an ablaut pattern CCaCC-*i* forming professional nouns (*ṣabun* ‘soap’ → *ṣwabn-i* ‘soap dealer’—note the nonlexical *w* filling the otherwise empty second C position); (b) MCA kin terms *ḥḥa* ‘father’, *ṃṃ* ‘mother’, and *xu-* (or *xa-*) ‘brother’; and (c) Spanish *padre* ‘father’, *madre* ‘mother’ (cf. *hermano* ‘brother’). First, the MCA and Spanish terms for ‘father’ and ‘mother’ were playfully blended to form *pp(ʷ)adri* and *ṃṃʷadri*, which are still in existence (but rare, and not known to many Moroccans). These can be analyzed phonologically as /pwaḍr-i/ and /mwaḍr-i/, respectively, since initial /pw/ and /mw/ in ablaut derivatives are pronounced [p̣p̣ʷ] and [ṃṃʷ]. The resulting forms are consistent in form with, and were probably influenced by, the CCaCC-*i* professional noun shape. In comparison with the corresponding MCA kin terms, the pattern can be taken more specifically as CCaḍr-*i*. Apparently the pharyngealization of the *ḍr* was taken as secondary (i.e., as having spread from the pharyngealized pronunciation of the initial labial clusters); alternatively, there may have been other (unrecorded) variants **pp(ʷ)adri* and **ṃṃʷadri* with unpharyngealized *dr*. The resulting CCaḍr-*i* pattern was then extended to ‘brother’ (MCA *xu-* or *xa-*), giving *xwadr-i*, which has gone on to have a life of its own even as *pp(ʷ)adri* and *ṃṃʷadri* have all but disappeared.

A form like *kīr-an* ‘intercity buses’ reflects a complex but linguistically normal and entirely unconscious set of structural forces. A form like *xwadr-i* must have originated as a brilliant linguistic joke, among a group of young men whiling away a muggy August afternoon in a street cafe. Or perhaps it was in a certain tavern in wartime Casablanca ...

References

- Heath, Jeffrey. 1987. *Ablaut and Ambiguity: Phonology of a Moroccan Arabic Dialect*. Albany: State University of New York Press.
- . 1989. *From Code-Switching to Borrowing: Foreign and Diglossic Mixing in Moroccan Arabic* (Library of Arabic Linguistics 9). London and New York: Kegan Paul International.
- . 1991. "Autour des réseaux dialectaux dans l'arabe des Juifs et des Musulmans marocains," in Issachar Ben-Ami, ed., *Recherches sur la culture des Juifs d'Afrique du Nord*, pp. xlix–lvi. Jerusalem: Communauté Israélite Nord-Africaine.

This page intentionally left blank.

Cypriot Arabic Phonology

Alexander Borg

Ben-Gurion University of the Negev

15.1. Introduction

Non-standard varieties of colloquial Arabic have attracted increasing interest among linguists during the last decades. Nonetheless, peripheral vernaculars of Arabic, i.e., those spoken by non-Arabs and/or outside the Arabic-speaking countries—for instance, in southeast Anatolia (cf. Daragözü, Jastrow 1973), central Asia (Fischer 1961; Sirat and Knudsen 1973; Kieffer 1980), parts of central Africa (Roth-Laly 1969; Hagège 1973), etc.—have yet to become the object of systematic individual study. This state of affairs is largely attributable to the fact that, despite their extensive formal deviance from normative Arabic, usually entailing a break in the link of mutual intelligibility between them and native varieties of Arabic, these vernaculars generally tend to be studied almost exclusively within the narrow ambit of Arabic dialectology. With the notable exception of Maltese (see chapter 16), only rarely have these erstwhile Arabic vernaculars been examined as autonomous linguistic systems pertaining to what are today, in effect, independent languages rather than merely fossilized relics of colloquial Arabic.

Given the further fact that most of these vernaculars survive today (a) in a highly undeveloped (when not terminal) state alongside major languages with long-established literary traditions; (b) as in-languages of specific ethnic groups; and, not infrequently, (c) in situations of complex interaction with other linguistic systems, the extent of their systemic autonomy (i.e., in synchronic terms) from surrounding languages would seem to be an issue that merits individual attention.

A particularly striking instance of such a linguistic situation exists among the ethnic *ſarap* of Afghanistan (Barth 1969: 11), whose special Arabic vernacular has developed in close contact with a variety of neighboring languages. Concerning their multilingualism, Kieffer (1980: 182) notes that “tous les hommes sont plurilingues: arabe, persan, ozbaki, torkman et, dans une certaine mesure, paštō,” and that they all speak Persian. A systematic

description of the differential synchronic impact exercised by these languages on the Arabic of Afghanistan would no doubt yield a highly complex pattern of linguistic interference. A preliminary sketch outlining some general trends within this complex pattern of linguistic interaction in Afghani Arabic has been recently undertaken in Ingham 1994.

The focus of the present study is a specific contact-related problem within the phonology of Cy[priot] A[rabic]—a special vernacular spoken exclusively by the bilingual (Arabic/Greek) Maronite villagers of Kormakiti (northwest Cyprus). Having developed since the Middle Ages in close proximity to Cy[priot] G[reek] and in complete isolation from the normative influences of vernacular and literary Arabic, this erstwhile Arabic vernacular today represents one of the most evolved offshoots of peripheral Arabic described to date.

The special evolutionary path taken by CyA vis-à-vis colloquial Arabic—paralleled *mutatis mutandis* in some of the aforementioned varieties of peripheral Arabic (particularly Maltese)—has left fairly conspicuous formal traces in its present structure: in particular, a considerable degree of systemic convergence with CyG in the realms of lexicon, syntax, and especially, phonology. The Greek specialist Brian Newton, who first showed an interest in the Greek adstrate in CyA, characterized this vernacular as a “mixed language” (Newton 1964; cf. also the designation “*morfosintassi mista*” in Lüdtke 1977), not least on account of its highly aberrant phonology.

The early history of Arabic in Cyprus is somewhat shrouded in obscurity; however, the emigration of Arabic speakers (“Syrians”) to Cyprus from the adjoining mainland at the end of the 12th century in response to an appeal by Guy de Lusignan is a well-attested fact (Jacoby 1979: 167). There is good reason to suppose that by this time CyA had already achieved a fair degree of formal distinctiveness from other varieties of Eastern Arabic. Thus though CyA has regrettably been almost completely overlooked by Western linguists and Orientalists until the middle of the present century, a reference to the existence of a Cypriot variety of Arabic occurs in a 13th century Arabic source: the *Kita:b as-Sima:t fi: asma:? an-naba:t* by ‘Izz ad-Dīn Abū Ishāq Ibrāhīm b. Muḥammad b. Tarxān as-Suwaydī (600/1204–690/1292). Summing up the contents of this work, Ullmann (1970: 291) makes the following observation, in which he mentions a number of areal varieties of Arabic recognized in as-Suwaydī’s time, including Cypriot Arabic:

As-Suwaydī gibt neben den arabischen Namen griechische (*yūnānī* und *rūmī*), syrische, persische, kastilische (*laṭīnī*) und berberische Bezeich-

nungen an. Bei den arabischen unterscheidet er zum Teil noch die Gebrauchsweise in Ägypten, Palästina, Spanien und Cypern.

Jacoby (1979: 168) has also drawn attention to the widespread use of Arabic in 14th century Cyprus (though apparently not specifically associated with the Maronite community) reported in a travel account compiled by the Augustinian monk Frater Jacobus de Verona (1335)—edited in Röhricht (1895)—which states

omnes de Cypro loquuntur grecum, bene tamen sciunt saracenicum et linguam francigenam, sed plus utuntur lingua greca. (p. 178)

‘all Cypriots speak Greek; however, they know Arabic and French well, but they use Greek more.’

In our own century, a rare mention of CyA occurs in Storrs and O’Brien (1930: 41), where it is designated as “a bastard Arabic mixed with Cypriot Greek” spoken “by the Maronites.” This lead, however, does not appear to have been taken up by subsequent research.

In Borg (1985: *passim*) I endeavored to show, among other things, that despite its *Sprachbund* relationship with the Greek-speaking area, CyA has, nonetheless, retained a distinctly Arabic typology inherited from a medieval Arabic colloquial displaying well-defined genetic affiliations with two fairly discrete dialect areas within contemporary Eastern Arabic:

(i) The *qētel*-group, comprising certain old sedentary Arabic dialects of north Syria and Mesopotamia which, like CyA, show the highly diagnostic Umlaut variety of the *imāla* shift (typified in Aleppine, and the *qaltu* dialects of Iraq and southeast Anatolia: CyA *klep* ‘dogs’ < OA *kila:b*, but CyA *nam* ‘he slept’ < OA *na:m*; for further details, cf. Borg 1985: 55–67; Blanc 1964: 42f.; Levin 1971; Jastrow 1978, *passim*).

(ii) The sedentary vernaculars spoken along the littoral of “Greater Syria” (e.g., Lebanese and Palestinian), with which it shares, for instance, the indicative verbal prefix {p-} < **b-*: /pisur/ ‘he becomes’.

Thus though the historical process that yielded the present CyA sound system has been extensively activated by language contact with CyG, comparative study of CyA phonology (cf. Borg 1985: 11–74) shows that the resulting phonological system still displays numerous regular sound correspondences with Arabic—a circumstance generally indicative of gradual historical evolution in CyA rather than sudden or catastrophic change.

Nevertheless, the present chapter is less concerned with outlining the systemic regularities that CyA shares with its erstwhile congeners on the

adjacent mainland than with pinpointing an area of “partial chaos” within the organization of the CyA sound pattern, specifically in the treatment of voicing in plosives. Before addressing the matter of voicing in CyA stops, it may be useful here to provide a general sketch of the CyA sound system so as to contextualize the data under examination and to account for certain trends untypical of Arabic in the surface treatment of CyA consonants.

15.2. The Cypriot Arabic sound system

The sound pattern of CyA presents the Arabic dialectologist with some of the most intriguing analytical problems in its entire structural profile. Not surprisingly, it is also in this domain of CyA that it shows the most profound systemic deviations from mainstream colloquial Arabic.

The phonological prime selected for the present analysis of the CyA sound pattern is a segmental unit closely approximating the traditional phoneme. This yields the inventory of consonants and vowels shown in Table 15-1.

Table 15-1. The Cypriot Arabic Segmental Paradigm (after Borg 1985)

Consonants					
p			t		k
	f	θ	s	š	x
	v	ð	z	ž	ʕ
m			n		
			l		
			r		
				y	
Vowels					
Monophthongs:	i		e		u
				o	
				a	
Diphthongs:	ay				aw

As can be seen from this display, the sound classes of CyA present a neatly symmetrical arrangement that is more akin to the contemporary CyG sound

pattern than to that of the medieval Arabic vernacular from which it evolved. Highly significant in this respect is the complete absence of an emphatic series in CyA, the O[ld] A[rabic] velarized consonants having been fused with their plain counterparts:

CyA		OA
<i>pasal</i>	‘onions’	<i>baṣal</i>
<i>kasel</i>	‘laziness’	<i>kasal</i>
<i>katt</i>	‘cat’	<i>qiṭṭ</i>
<i>fatt</i>	‘it was slight in quantity’	<i>fatt</i>
<i>peḏa</i>	‘egg’	<i>bayḏa</i>
<i>iḏn</i>	‘ear’	<i>uḏn</i>

The extensive impact exerted by CyG on the CyA sound system should be evident from the sketch of CyA phonology provided in Borg (1985). In effect, the systemic encroachment of CyG on CyA amounts to more than a mere case of historical language contact; it represents rather a progressive (and still ongoing) hellenization process. The potential impact of CyA on the Greek spoken by the monolingual Maronites (from Aghía Marína, Asómatos, and Karpásha) has yet to be investigated.

Highly symptomatic of the profound inroads made by Greek into CyA sound structure is the extreme contraction of the Arabic segmental paradigm actualized in the latter. Thus in contrast with the 35-odd segments typifying the sound systems of most sedentary dialects of Eastern Arabic (cf. Damascus, Cantineau and Helbaoui 1953: 14–15; Tripoli, El-Hajjé 1954: 11f.; Cairo, Selim 1967: 135; etc.), CyA has—according to my 1985 analysis—only 26 segments, a figure that closely matches the 25 segments of CyG (Newton 1972a: 22). As can be seen from the schematic display above, the selection of articulatory classes in CyA is closely modeled on that of CyG; thus, only one point of articulation, that of the CyA voiced pharyngeal fricative /ʕ/, is completely foreign to CyG. In the other direction, the CyG segments /ɕ/ and /kʰ/ (palatalized velar stop) do not form part of the CyA segmental inventory, but are in essence present in CyA surface structure as the biphonemic sequences /tɕ/ and /ky/, respectively. The cumulative impact of these systemic shifts has not only rendered the surface content of CyA and CyG highly similar but, as will be seen below, it has also occasioned a notable historical restructuring of the function of voicing in obstruents; the descriptive details and wider implications relating to this change will be dealt with at length in the main section of this chapter.

15.3. Some phonological rules in Cypriot Arabic

15.3.1. Manner dissimilation

As in Cypriot and Standard Greek, biconsonantal stop clusters in CyA are subject to a manner dissimilation constraint (both diachronic and synchronic) replacing the first stop by its corresponding fricative (cf. Gk. /nixta/ < *nikta* ‘night’):

/xtilt/	‘you (masc.sg.) killed’	← ktilt	< * <i>qtilt</i>
/fkum/	‘I get up’	← pkuml	< * <i>baqu:m</i>

The unmarked character of “fricative + stop” CC clusters in CyA occasionally also yields secondary occlusivization of underlying or historical spirants: /xtir/ ‘much’ < **kθi:r*.

15.3.2. Palatalization and occlusivization of /y/

The CyA velar segments /k/ and /x/ undergo automatic palatalization (here indicated by an apostrophe) before the historical and/or underlying front vowels /i/ and /e/ and the palatal glide /y/:

[k'ɪlp]	‘dog’	< <i>kalb</i>
[k'el]	‘he ate’	< * <i>ke:l</i> < <i>akal</i>
[x'ɪrep]	‘he left’	< * <i>hirib</i>
[x'érep]	‘leaving’	< <i>ha:rib</i>
[x'ar]	‘cucumbers’	< <i>xya:r</i>
[pk'ut]	‘houses’	< <i>byu:t</i>

The secondary velar stop in the last example is the product of a phonetic tendency in CyA (ultimately carried over from CyG) towards inserting an epenthetic velar stop [k] in dyadic clusters consisting of obstruent + palatal glide:

/pkɪyara/	‘wells’	< * <i>bya:ra</i>
/apkyað/	‘white (masc.sg.)’	< <i>abyaɖ</i>
/θkyep/	‘clothes’	< * <i>θye:b</i>

In certain contexts, this secondary velar stop is sometimes liable to partial morphologization, as in /mišku/ ‘they walked’ (< **miškyu* < **mišyu*), and /ʕosk/ ‘sticks’ (< **ʕosky* < *ʕuʕiyy*), where the velar stop can be said to have replaced etymological /y/ as a third radical.

However, complete assimilation of the alveolar nasal and of the lateral liquid to a following /y/ is not an automatic process in CyA but appears to be restricted to certain morphological boundaries or to specific lexical items: /kayyisrok/ ‘he used to steal’ < **ka:n yisroq*, /teyye/ ‘vine’ < **da:lya*, etc.

15.3.3. *Postnasal epenthesis*

A fairly automatic phonetic rule in CyA requires the insertion of an auxiliary stop segment between a nasal and a following continuant or resonant, the secondary plosive agreeing with the preceding nasal in point of articulation and with the following continuant in voicing: [intsàn] ‘man, husband’ < *in-sa:n*, [kampx] ‘wheat’ < *qamḥ*, [x’imbl] ‘load’ < *ḥiml*, etc. Though no comparable sound rule has been noted for Cypriot Greek dialects in Newton (1972a, 1972b), this epenthesis rule in CyA appears to have been contracted from Greek (Johannes Niehoff, personal communication 1994).

15.4. Stop voicing in Cypriot Arabic

Stop voicing does not ordinarily present a problem area in descriptive work on mainstream varieties of spoken Arabic, where this contrast continues a formal trait inherited from OA, and ultimately, from older forms of Semitic. In Borg (1985: 12–16), I noted that special features relating to obstruent voicing in CyA arise principally from this vernacular’s history of language contact with CyG, which has culminated in the high degree of Arabic/Greek bilingualism characterizing the contemporary Kormakiti community.

A concise but fairly detailed account of the surface distribution of obstruent voicing in CyA was presented in Borg (*ibid.*), where it was concluded that notwithstanding occasional erratic reflexes of OA obstruent voicing in CyA surface structure, the functional role of stop voicing in CyA is too negligible to be systemically significant. My purpose in resuming the discussion of this topic here is to react to critical observations that appeared in two reviews of my monograph: Hopkins (1990: 89) and Singer (1991: 408). Both scholars have suggested that recognition of a voicing opposition in CyA stops could conceivably yield a more insightful phonological analysis of the vernacular than that proposed in my aforementioned study, where this feature is assigned merely allophonic status.

In the following sections I propose to sketch some salient diachronic and synchronic trends relevant to CyA voicing and then to examine some emendations to my phonological analysis proposed in Hopkins’ critique of the voicing issue, in order to determine what if anything in my original presentation requires modification.

One final remark here relates to Hopkins’ view (p. 89)—echoed by Singer—alleging that my phonological analysis of CyA is “rather complicated.” Hopkins, however, also recognizes the intricate character of CyA phonology

when contrasted with the transparency of most other colloquial Arabic sound systems. Thus no mainstream Arabic dialect investigated so far displays a comparable degree of complexity in its phonetic treatment of underlying stop segments. As I noted in Borg (1985: 6), the complexity of CyA phonology derives simultaneously from its alignment with the CyG sound system in the realm of phonotactics and from its obvious lack of isomorphy with CyG in the domain of root structure (CyA and CyG being genetically unrelated languages). Specifically, while root morphemes in Greek, as in other Indo-European languages, comprise continuous strings displaying both consonantal and vocalic components (cf. Schwyzer 1938: 419), CyA word structure has retained the carried over mobility of Arabic radicals, with the result that phonological rules contracted from CyG, like manner dissimilation, post-nasal voicing, palatalization, etc., represent essentially historical shifts in Greek but synchronically productive processes when carried over into CyA. One example will suffice here. The aforementioned manner dissimilation rule requires the first segment in Ancient Greek stop + stop clusters to undergo spirantization in many varieties of Greek, including CyG; cf. Ancient Greek $\pi\tau\omega\chi\acute{o}\varsigma$ 'poor', $\delta\acute{o}\kappa\tau\acute{o}$ 'eight' > CyG *ftoxòs*, *oxtò*, etc.). Observe now the morphophonemic spirantization of the velar and labial stops before /t/ in the CyA root {k-t-p} (< OA *kataba* 'he wrote'):

CyA

<i>kitep</i>	'he wrote'	⇐ kitipl
<i>xtuft</i>	'I wrote'	⇐ ktuptl
<i>paxtop</i>	'I write'	⇐ paktopl

As will be seen below, the endeavor to maintain invariant underlying representations for roots in a synchronic account of the CyA sound system introduces a certain degree of complexity into the phonological analysis of voicing in this vernacular.

15.4.1. Paradigmatic aspects of voicing in Cypriot Arabic

Voicing fulfilled a purely marginal role in the OA stop series, since, in effect, only dentals implemented this phonological opposition:

–	t	k	q	?
b	d	–	–	–

Its functional load was greater in the OA fricative series, where four of the seven available articulatory positions within this class were occupied by voiced/voiceless pairs:

f	θ	s	š	x	ħ	h
—	ð	z	—	ɣ	ʕ	—

The majority of modern Arabic dialects have generally extended and evened out the paradigmatic domain of the voicing opposition within their obstruent systems, usually by filling out the Old Arabic “empty slots” through internal change and/or integration of foreign sounds, e.g., Baghdadi Arabic (Blanc 1964: 17), the Jewish dialect of Algiers (Cohen 1912: 520–21), etc.

As shown in Table 15-1 above, CyA has, like most Arabic dialects, also developed a high degree of symmetry in its paradigmatic implementation of the voicing opposition. The impact of the CyA sound system on that of CyG has entailed for CyA both positive and negative trends, promoting internal symmetry and formal alignment with Greek in the functional contrasts implemented.

On the positive side, CyA has extensively retained the interdental articulation of *θ and *ð (< OA *ṯ and *ḏ) lost in many sedentary—principally urban—varieties of Eastern Arabic, and OA /w/ has shifted to CyA /v/; in both cases, the systemic role of voicing among CyA fricatives has been enhanced. The retention in CyA of a voiced pharyngeal fricative /ʕ/ (without a voiceless counterpart) is highly noteworthy given its complete extraneousness to the CyG consonant inventory; its survival in CyA would seem to suggest a relatively late date for intensive contact between CyA and Greek.

On the negative side, the CyA consonant paradigm shows segmental streamlining through the merger of a number of OA contrasts:

OA		CyA
ṣ, s	>	s
ṣ̣, ḏ	>	ð
ħ, x, h	>	x
ʕ, ɣ	>	ʕ

The absence of the voiced uvular fricative *ɣ in CyA is striking given the presence of CyG /ɣ/. This latter factor would seem to suggest that the loss of this segment was a substratally inherited trait antedating linguistic contact with Greek.

From the specific standpoint of the voicing feature, the systemic contraction of the CyA consonant paradigm resulting from the merger of these OA fricatives has been compensated for by the aforementioned shift of *w > /v/, and of *ǧ > /ʒ/, both of which have voiceless counterparts in the CyA paradigm. The CyA velar fricative /x/ lacks a full-fledged voiced correlate in the inherited lexicon, but an intriguing CyA phone [ɣ] tends to show up

in surface reflexes of certain secondary clusters: Ar. **kb*, **kd* > **gb*, **gd* > CyA [ɣb], [ɣd], respectively: CyA [ɣbiːr] ‘big (masc.)’ < *gbiːr* < Ar. *kbiːr*, [móɣde] ‘stove’ < **moːqde* < OA *mawqida*, etc. In fact, as we shall see (in § 15.5.3), the formal interpretation of these voiced clusters in the synchronic phonology of CyA is not easily accommodated within the scope of a conventional phonological analysis.

In addition to its elimination of the voicing contrast in stops, formal restructuring within the CyA stop system has also entailed a reduction in the number of articulatory classes retained. Significantly, these latter comprise the labial, dental, and velar points of articulation—a selection that was partly dictated by the paradigmatic structure of the CyG sound system and partly by a more general naturalness factor, these being the three phonologically least marked articulatory positions along the vocal tract. (Note, for instance, the loss of the highly marked uvular position via fusion of OA **q* and **k* > CyA *k*.)

Thus, like the Greek dialect family, CyA displays the “classic” stop series, /p/, /t/, and /k/; these are voice-indifferent and derive historically in the following fashion:

CyA		OA
<i>p</i>	<	<i>b</i>
<i>t</i>	<	<i>t, ʈ, d</i>
<i>k</i>	<	<i>k, q</i>

Examples:

CyA		OA
<i>pirek</i>	‘it flashed (lightning)’	< <i>baraq</i>
<i>tute</i>	‘mulberry tree’	< <i>tuːta</i>
<i>tute</i>	‘worm’	< <i>duːda</i>
<i>tar</i>	‘he turned round’	< <i>daːr</i>
<i>tar</i>	‘it flew’	< <i>ʈaːr</i>
<i>kiser</i>	‘he broke’	< <i>kasar</i>
<i>kitel</i>	‘he killed’	< <i>qatal</i>

The pairs of CyA homophones with dental stops provide a striking illustration of the historical loss of the voicing opposition in this vernacular.

15.4.2. Syntagmatic features of Cypriot Arabic stops

CyA utterances display both voiced and voiceless stops. As already noted in Borg (1985: 12), voiced simplex (i.e., non-geminated) stops occur most often

in resonant phonetic contexts, i.e., intervocalically, and in contact with the resonant segments [m], [n], [l], and [r], or in the adjacency of /z/:

[síbel]	‘stubble’	/sipel/
[k’ídel]	‘he killed’	/kitel/
[tága]	‘window’	/taka/
[xágle]	‘field’	/xakle/
[págra]	‘cow’	/pakra/
[índi]	‘you (fem.sg.)’	/inti/
[ʕúrba]	‘strangers’	/ʕurpa/
[várde]	‘flower, rose’	/varte/
[varga]	‘book’	/varka/

In contact with the voiced alveolar spirant [z], CyA stops are always voiced:

[bzáfʔat]	‘I feared’	/pzaʔat/
[pázga]	‘spittle’	/pazka/
[nágza]	‘sharp pain’	/nakza/

Outside this specific environment, CyA stops attain full phonation optimally after nasals:

/np/	⇒	[mb]
/nt/	⇒	[nd]
/nk/	⇒	[ŋg]

In other contexts potentially conducive to stop voicing, such as intervocalically and in contact with liquids, it is often difficult to distinguish with certainty between voiced stops and lax, voiceless ones. Sound spectrograms taken of CyA utterances do not always show a distinctly discernible voice bar accompanying intervocalic simplex stops. However, previous researchers of CyA have also tended to perceive phonatory activity here (e.g., Roth 1975, *passim*).

Voiceless realizations of CyA stops are most common pre- and post-juncturally where voicelessness is the norm for stops even in contact with a following resonant, usually with concomitant devoicing of the latter:

[pṛúde]	‘coolness’	/prute/
[kḷep]	‘dogs’	/klep/
[kapṛ]	‘grave’	/kapr/
[tipṇ]	‘straw’	/tipn/

Like CyG, CyA retains historical stop gemination. The geminate clusters /pp/, /tt/, and /kk/ are always voiceless and highly aspirated; /dd/ is generally also voiceless but unaspirated. The unsymmetrical relationship between CyA simplex stops and their corresponding geminates is discussed below.

15.5. Discussion

To sum up the preceding section, laryngeal features in surface realizations of CyA labial, dental, and velar stops are generally predictable from immediate phonetic environment. Simplex CyA stops are realized:

- (a) voiceless, lax, and unaspirated in word-initial and -final positions;
- (b) lax, (usually) voiced intervocalically and after resonants, particularly nasals;
- (c) voiced in contact with /z/.

Geminated labial and velar stops are invariably tense, voiceless, and aspirated, whereas geminated dental stops can be either tense, voiceless, and aspirated; or lax, voiceless, and unaspirated.

The non-functional status of stop voicing in CyA was, incidentally, also independently postulated in Jastrow (1977) on the basis of the notational vagaries and inconsistencies in Tsiapera's (1969) descriptive sketch of this language. Arlette Roth's 1975 study of CyA verb morphology also assumes diachronic systemic loss of the voicing feature among stops in this vernacular.

As we shall see below, in the few cases where a voice contrast in stops may be ostensibly relevant from the perspective of a conventional segmental analysis, there are often good reasons for questioning the validity of this solution.

In the following sections, I shall take issue with Hopkins' argumentation in favor of a voicing opposition in CyA stops. In effect, Hopkins recognizes three sources of a potential voicing contrast in this vernacular:

- (a) post-nasal stop voicing;
- (b) the opposition of the CyA geminates /tt/ : /dd/; and
- (c) residual retention of Old Arabic voicing in certain dyadic clusters: CyA [ɣbi:r] 'big' < Ar. *[gbi:r] < *kbi:r*. Let us here consider each of these points in turn.

15.5.1. *Post-nasal voicing*

It should be clear from the preceding sections that the oppositions *p* : *b* and *k* : *g* never materialize in unambiguous phonological contexts, e.g., intervocalically. That CyA reflexes of historical labial and velar stops should be voice-indifferent is hardly surprising since, as already noted, there was no historical source for these oppositions in OA. Sedentary Arabic vernaculars which have functionalized these potential contrasts have generally achieved this through the integration of lexical loans with [p] and [g]: Baghdadi /pu:ʃi/ 'veil', /parda/ 'curtain' (< Persian); for Damascene /g/, see Ferguson (1969).

By way of contrast, CyA appears to have been fairly well shielded from contact with foreign lexical influences other than those filtered through CyG.

Surface [b], [d], and [g], however, do sometimes occur post-juncturally in native CyA words, mostly verbs; here the voiced segment invariably derives from underlying bi-phonemic sequences consisting of nasal + stop, i.e., /np/, /nt/, and /nk/, respectively. Such word-initial sequences are, for instance, particularly frequent in verbs displaying the passive marker {N-}:

[pálas]	‘he ran over’	
[mbálas ~ bálas]	‘he got run over’	← N + palas
[tílef]	‘it leaked’	
[ndílef ~ dílef]	‘it got soaked’	← N + tilef
[k’íser]	‘he broke’	
[ŋg’íser ~ gíser]	‘it got broken’	← N + kiser

Hopkins, however, contends that the initial voiced stops in the optional variants with a deleted nasal furnish a sufficient basis for setting up voiced stop phonemes in CyA:

Borg is doubtless right in principle in saying that simple *d* and *t* no longer contrast, but there do seem to be cases where this opposition is possible. One of them is (p. 13) *tilef* “it leaked” : *dílef* “it got soaked” (~ *ndílef*). The fact that *dílef* has an alternative form should perhaps not be made to carry more than it can bear: if the form *dílef* exists and a speaker uses it, we have ipso facto a phonemic contrast between *d* and *t*. Borg himself says as much when he calls this opposition “functionally significant.” (p. 89)

Hopkins’ conventional reasoning is somewhat off the mark. An insightful appraisal of surface voicing must not overlook the fact that post-nasal stop voicing (with or without nasal deletion) is a quintessential hallmark of Greek phonic interference. I recall here some remarks from Hamp (1962: 641) relating to the perception of stop voicing by speakers of Greek, where, incidentally, the same analytical problem (i.e., the phonological status of voiced stops) occurs:

One can judge from various aspects of the reactions of Greek speakers that they hear voiced stops and affricates in typical European languages as being nasal, because, for all Greek except acculturated varieties often heard in Athens, a voiced phone occurs either (i) automatically preceded by a nasal segment, or (ii) *in free variation with nasal plus voiced stop segments* ... (Hamp 1962: 641, emphasis added)

The intimate link between nasals and stop voicing in CyA can be seen in its diachronic treatment of stops, e.g., the occasional shift of Ar. **dd* > CyA [nd]:

CyA	OA
[k'indám] 'before'	< <i>qudda:m</i>
[šánda] 'he entered'	< <i>šadda:</i>

If our synchronic analysis of CyA is to reflect the speaker's perception of the underlying facts, the phonological interpretation of the voiced initial stops here must surely be /n/ + voiceless stop.

Even in the event that CyA speakers were to lose the longer variants with {N-} cited above and to retain the counterparts with the initially voiced stops, there are reasons of a morphological nature why one would not want to analyze these initial voiced stops as unit phonemes. Positing the voiced stops /b/, /d/, and /g/ as underlying segments in this grammatical context would be tantamount to setting up phonological segments that harbor morphological boundaries. One would, in other words, be postulating a CyA "reflexive marker" with the odd realization [+voice]. Surely both of these features are inadmissible in a conventional linguistic analysis, particularly in the case of the CyA passive marker {N-}, which has been retained as such in phonetic environments that do not require its fusion with the stem-initial consonant:

<i>šarat</i>	'he ripped'	
<i>ntšarat</i>	'it got ripped'	⇐ N + šarat
<i>faylet</i>	'he loosened'	
<i>mfaylet</i>	'it got loosened'	⇐ N + faylitl

The simplicity of the morphological analysis of CyA would thus be notably enhanced if we set up an invariant basic form of the passive morpheme {N-} which is liable to optional deletion before phonetically voiced stops thus:

<i>Underlying form</i>	N + tilef 'it got soaked'
<i>Post-nasal voicing</i>	[ndílef]
<i>Optional nasal deletion</i>	[dílef]

15.5.2. The CyA opposition /t/ vs. /dd/

One of the special features of the CyA stop series is its differential historical treatment of single vs. geminate dentals. Whereas the present phonetic realization and distribution of CyA phones continuing the OA simplex dentals attest to the complete fusion of the OA voiced and voiceless dental series into CyA /t/, the vernacular nonetheless retains, to a certain extent, the distinction between reflexes of geminated OA /tt/ and /dd/. The phonetic basis of the latter contrast in CyA, however, is not the voicing feature, but a tense : lax opposition. Like other geminated stops, CyA /tt/ is uttered without pho-

nation and with fairly strong aspiration; CyA /dd/, on the other hand, is voice-indifferent (generally voiceless) but never aspirated.

In fact, only a few permanently voiced reflexes of historical or secondary Ar. **dd* have been retained in CyA; as already noted in the previous subsection, these tend to show the shift Ar. **dd* > CyA *nd*, e.g., [k'indám] 'before' < Ar. *qudda:m*, etc. The special historical treatment of dental segments in CyA has yielded the following unusual patterning of phonological oppositions among plosives (the dash represents an empty paradigmatic slot for a voiced stop):

Single stop	Geminated stop
[p] : –	[pp] : –
[t] : –	[tt] : [dd]
[k] : –	[kk] : –

The phonetic bases of this three-way contrast among dentals can be tabulated as in Table 15-2:

Table 15-2. Cypriot Arabic Dental Stop Contrasts

	<i>t</i>	<i>tt</i>	<i>dd</i>
voice	±	–	±
aspiration	–	+	–
length	–	+	+

The oppositions between CyA dentals are optimally implemented in word-medial and -final positions:

CyA	Arabic
<i>ʕata</i> 'lid'	< <i>yīṭa:ʔ</i>
<i>ʕatta</i> 'he covered'	< <i>yattā:</i>
<i>ʕadda</i> 'he bit her'	< <i>ʕaḍḍha:</i>
<i>mat</i> 'he died'	< <i>ma:t</i>
<i>fatt</i> 'he stripped off (fruit)'	(~ ? <i>faraṭ</i>)
<i>madd</i> 'he measured'	< <i>madd</i>

In word-initial position, the contrast between simplex and geminate dentals falls outside the strictly phonological sphere, since initial geminates in the native CyA lexicon generally include a morphemic or lexical boundary and tend to cooccur with certain grammatical processes, e.g., marking for definiteness by means of the article /l/, which in CyA completely assimilates to most initial consonants:

<i>ddist</i>	‘the dish’	⇒ [t:ist]
<i>ddayn</i>	‘the debt’	⇒ [t:ayn]
<i>ttimm</i>	‘the blood’	⇒ [t ^h im]
<i>ttik</i>	‘the rooster’	⇒ [t ^h iːk]

The words just cited exemplify the analytical problem under discussion in a striking manner. Without the article, all four lexical items display an initial lax, voiceless, dental stop [t]:

[tist]	‘dish’	< Ar. <i>dist</i>
[tayn]	‘debt’	< Ar. <i>dayn</i>
[tim]	‘blood’	< Ar. <i>damm</i>
[tiːk]	‘rooster’	< Ar. <i>diːk</i>

Once the article /l-/ is preposed, a latent contrast between CyA dentals comes to light. Within the theoretical context of a conventional phonological analysis, this is an unusual situation, since morphophonemic processes (e.g., morphologically generated consonantal gemination) do not ordinarily yield functional contrasts unattested at the purely phonological level.

One could here recognize an abstract *t* : *d* contrast between simplex dental stops, whose phonetic correlates are always neutralized. Since phonological neutralization as customarily understood eliminates an opposition exclusively in specific contexts (e.g., prejunctural voicing in languages such as German, Turkish, Russian, Maltese, etc.), the setting up of such an opposition in CyA would entail admitting a case of complete neutralization.

It should be clear from the last two sets of examples that the lexical distribution of the CyA geminates /tt/ and /dd/ in this morphosyntactic context corresponds only partially to the historical voice opposition between Old Arabic **t* and **d*. In fact, in the greater number of cases, historical **d* tends to yield simplex /t/ and geminate /tt/, as in the following words:

CyA		Old Arabic	
<i>t-tarp</i>		‘the road’	< <i>d-darb</i>
<i>t-teyye</i>		‘the vine’	< <i>d-da:liya</i>
<i>t-taken</i>	< * <i>d-daqa</i> n	‘the beard’	< <i>ḏ-ḏaqan</i>
<i>t-tuppane</i>	< * <i>d-dubba:na</i>	‘the fly’	< <i>ḏ-ḏubba:na</i>

The synchronic distribution of these CyA reflexes is, incidentally, also highly subject to idiolectal variation—possibly reflecting the terminal state of the vernacular.

From the diachronic standpoint, the phonological contrast between CyA *tt* and *dd* represents the outcome of an intriguing fusional process drawing

simultaneously on both OA and CyG sound systems. Thus, while on the one hand, this functional opposition obviously continues in part the OA voicing contrast *tt* : *dd*, its phonetic implementation in CyA exploits the Greek tense : lax distinction (realized phonetically as aspirated vs. non-aspirated release) which in CyG, however, opposes simplex to geminate stops and never two different kinds of geminates. Thus the opposition *tt* : *dd* is completely foreign to CyG phonology, where only the following simplex : geminate contrasts obtain in the dental and interdental classes (Newton 1972a: 34):

CyG:	<i>t</i> : <i>tt</i>	<i>θ</i> : <i>θθ</i>
		<i>ð</i> : <i>ðð</i>

The *tt* : *dd* contrast in CyA amounts to an innovation in another sense too: no other Arabic vernacular described so far utilizes phonetic aspiration as a morphological device:

Basic stem	Derived stem
<i>šite</i> [šíde] 'rain'	<i>šatta</i> [šát ^h a] 'it rained'
<i>kapr</i> [kapr̥] 'grave'	<i>kapper</i> [káp ^h er] 'he buried'

The analytical problem relating to the relationship between simplex and geminated dentals also shows up in another context (not discussed in Hopkins): morphologically conditioned degemination of CyA /dd/, for instance, in broken plurals of certain CyA nouns whose plural form displays two occurrences of /t/, the first realized voiceless, the second showing intervocalic voicing. One could here adopt a more abstract phonological notation than the one used in Borg (1985) in order to distinguish between simplex dentals that yield /dd/ when geminated from those that yield /tt/. The former could be indicated by means of an abstract symbol *d* and the latter by means of *t*, as in the following two lexical sets:

<i>kiddás</i>	'Mass'	
<i>xdedís</i>	'Masses'	⇒ [xtedí's]
<i>fiddán</i>	'yoke of oxen'	
<i>fdedín</i>	'yokes of oxen'	⇒ [ftedí'n]
<i>štume</i>	'insult (n.)'	
<i>šattem</i>	'he insulted'	⇒ [šát ^h em]
<i>xtat</i>	'cats'	
<i>katt</i>	'cat'	⇒ [kat ^h]

In Borg (1985) I refrained from doing so since this would have meant recognizing a *t* : *d* contrast in simplex CyA dentals for which there are no stable phonetic correlates. This would have also had the effect of obscuring the

extent of historical change actualized in CyA. Above all, given the fact that voicing in simplex CyA plosives is conditioned by the immediate phonetic context, it seemed to me highly artificial to set up a genuine voicing contrast. Paradoxically, the recognition of an abstract *t* : *d* distinction (in place of the full-fledged *t* : *d* phonemic contrast suggested by Hopkins) would ideally have to ignore phonetic voicing as a criterion, since the distribution of voiced and voiceless stops in CyA is automatically predictable.

15.5.3. Voicing in dyadic clusters

A striking oddity in the synchronic behavior of CyA stops in certain words is their tendency to show up voiced (and spirantized in conformity with manner dissimilation) in immediate contact with other stops. Thus, for instance, the CyA root {k-p-r} ‘big’ yields the following surface forms, among others:

- | | |
|--------------|------------------|
| (a) [k'íber] | ‘he grew up’ |
| (b) [k'íbet] | ‘she grew up’ |
| (c) [ɣbiːr] | ‘big (masc.sg.)’ |
| (d) [ɣbar] | ‘big (pl.)’ |

Voicing of /p/ in the resonant environments (a) and (b) conforms to CyA voicing norms outlined in § 15.4.2; voicing of the word-initial clusters in (c) and (d), on the other hand, represents a residual reflex of regressive voicing assimilation in Ar. **gbiːr*, **gbar* < OA *kabiːr*, *kibar* that survived the paradigmatic loss of stop voicing in the CyA plosive class. Three voiced clusters of this kind commonly occur in CyA: [ɣb], [ɣd], and [ðb]. The dyadic clusters showing [ɣ] in first position seem to represent a “frozen” phonotactic template apparently unknown in CyG, where—assuming that the data on CyG consonant clusters provided in Newton (1972a: 29) are complete—/y/ clusters exclusively with resonants: [ɣl], [ɣr], [ɣn], [ɣm]. Highly indicative here is the metathesis of stops in CyA [ɣba] ‘he wept’ generating this voiced cluster through regressive voicing assimilation: *[bkaː] > *[kbaː] > *[gbaː] > CyA [ɣba]. Some of these clusters also occur in the morphophonemics of verb inflection, e.g., of /rikep/ ‘he rode’, /páxxer/ ‘he incensed’, /pas/ ‘he kissed’, /tápax/ ‘to cook’:

- | | |
|---------------|----------------------|
| [ríɣbet] | ‘she rode’ |
| [piðbáxxer] | ‘she incenses’ |
| [piðbús] | ‘she kisses’ |
| [pittíniðbex] | ‘it is being cooked’ |

In my monograph on CyA, I analyzed voicing in such clusters as subphonemic, despite the fact that, distributionally speaking, phonological contrasts with corresponding voiceless clusters do occur, as in the following contrastive pairs:

CyA		Arabic
[ɣbar]	‘big (pl.)’	<i>kiba:r</i>
[xpaz]	‘loaves’	<i>axba:z</i>
[ɣdi:ʃ]	‘horse’	<i>kadi:ʃ</i>
[xtiˈr]	‘much’	<i>kaθiˈr</i>
[móyde]	‘stove’	<i>mawqida</i>
[oxt]	‘time’	<i>waqt</i>

which I transcribed phonologically: /xpar/, /xpaz/, /xtiʃ/, /xtil/, /moxte/, and /oxt/, respectively. Hopkins (1990: 89–90) objected to my phonological representations of the voiced clusters:

The analysis of forms such as /xpɪr/ ‘big’ with obligatory voicing [ɣbiˈr] seems to me unnecessarily complicated and abstract. If voicing of /xp/ is here obligatory, then one should think of adding either *ɣ* or *b* (or both) to the phonemic inventory. Borg, however, is unwilling to do this since he does not admit a phonemic opposition *b* : *p*, *d* : *t*, etc., and proposes instead to solve the problem by employing the ad hoc compromise device of underlining these combinations: /xp/ = [ɣb]. This is unconvincing. When Borg avers (p. 16) that the “distribution of voicing in these dyadic clusters is synchronically unpredictable” this is tantamount to saying that the proposed phonemic analysis cannot account for it.

Hopkins here adopts the viewpoint of traditional phonemics according to which phones occurring in contrasting environments are accorded systemic status. However, he overlooks other formal factors that merit consideration and which suggest, at the very least, that the CyA consonant paradigm has a gray area in its assignment of voicing:

a) CyA has extensively retained the root and scheme composition of stems typical of Arabic and other Semitic languages. This factor makes it desirable that the phonological analysis posit wherever possible invariant underlying representations for CyA root morphemes. Given the circumstance that CyA speakers perceive lexical items such as [rígep] ‘he rode’, [rkift] ‘I rode’, and [ríybu] ‘they rode’ as semantic cognates, the phonological analysis should reflect these native perceptions and assign these items identical radicals. This procedure is particularly desirable since this type of spontaneous voicing in CyA clusters is also liable to show up in morphological inflection, e.g., in the following verb forms:

CyA		Arabic
[sak ^h ep] /sakkep/	‘he overtook’	< <i>sabbaq</i>
[saybet] /saxpet/	‘she overtook’	< <i>sabbaqat</i>
[kʰidep] /kitep/	‘he wrote’	< * <i>kitib</i>
[kʰiðbet] /kiθpet/	‘she wrote’	< * <i>kitbit</i>

Recognition of /b/, /d/, and /g/ as phonemic segments would necessarily imply an allomorphic split of these verbal roots thus: {s-k-p ~ s-y-b}, {k-t-p ~ k-ð-b}, and so on throughout the CyA lexicon. The artificial nature of this “solution” can easily be inferred from the occurrence of free variants such as [rákifti] ~ [raybíti] ‘my neck’ < OA *raqabati*;, which attest to the relative automaticity and low-level status of the voiced clusters under discussion. Hopkins refers (p. 90) to my diacritical underlining of these spontaneously voiced clusters as a “cosmetic” device. However, another formal interpretation is also possible. Given the nonexistence of independently voiced segments in CyA outside the context of these dyadic consonant clusters, one could simply mark the roots that are affected by this idiosyncratic trend in preference to setting up full-fledged voiced plosives in the phonemic inventory. This treats spontaneous stop voicing in clusters as a feature pertaining to certain lexemes: in other words, a lexical rather than a purely segmental feature.

b) It is significant that the combinatory potential of these voiced segments is also extremely restricted. The aforementioned treatment of OA *baka*: (e.g., the metathesis affecting the etymological plosives **bk*) is particularly indicative in this respect. Interestingly, underlying or historical dyadic obstruent clusters with [b] in first position do not seem to occur in CyA; observe, in this connection, the regular devoicing of historical **bd* in CyA /ptilt/ ‘I changed’ < **bdilt* < OA **badalt*. Significantly CyA here also retains the non-spirantized surface stop [p], yielding a minimal contrast with /ftilt/ ‘I twisted’.

15.6. Synopsis and conclusion

It should be evident from this presentation that no strong case for postulating a systematic voicing contrast in underlying representations of CyA stops can be sustained on the basis of the data at hand. As has been noted, voicing of CyA stops occurs in fully or partially conditioning environments, i.e., after nasals, and, less predictably, in certain dyadic clusters which undergo voicing in certain morphophonemic contexts. In both cases, the potential functional role of the voicing feature in stops is anyway fairly negligible. The

case of CyA *tt* : *dd* is also of marginal relevance to the issue of voicing, since neither sequence is in fact intrinsically and permanently voiced. The only other clear case of regular surface voicing affecting underlying CyA stops occurs in the clusters examined in the preceding section, where the cooccurrence of voicing and clustering makes an unambiguous appraisal of the functional role of voicing per se somewhat problematic.

I would like to suggest that the formal ambiguity obtaining vis-à-vis the functional role of the CyA voicing feature here is comparable to that of consonant length in Modern Hebrew. Observe, for instance, the following modern reflexes of Biblical Hebrew *ḥ-g-g* ‘Fest’ (Buhl 1954: 213; cf. Ar. *ḥaġġ* ‘pilgrimage’), where etymological geminated **g* is simplified in items (a) and (b) but retained in (c), (d), and (e):

Modern Hebrew

- | | |
|-------------------|----------------|
| (a) <i>ḥag</i> | ‘feast’ |
| (b) <i>ḥag-ím</i> | ‘feasts’ |
| (c) <i>laḥgóg</i> | ‘to celebrate’ |
| (d) <i>ḥagigá</i> | ‘celebration’ |
| (e) <i>ḥagigí</i> | ‘festive’ |

Though native Hebrew speakers undoubtedly perceive a clear semantic and derivational relationship between all these lexical items, functional loss of consonant length in Modern Hebrew has rendered the link between their surface and underlying form less transparent than it was for speakers of the ancient language.

In their reviews of Borg (1985), Hopkins (1990) and Singer (1991) argue for a “voicing” contrast in CyA, assuming too readily (despite indications to the contrary in Borg 1985: 12–16) that a stable correlation between phonetic and phonological voicing actually exists. In this they may have been misled by my use of the letters *p t k b d g*. Following traditional practice in phonological descriptions, I resorted to these symbols to characterize more than two different laryngeal states in phones and phonemes uttered in the labial, dental, and velar regions of the vocal tract, the “voiceless” set [p t k] being always assumed to represent the unmarked phonetic value within each allophonic split (voiceless : voiced; tense : lax; aspirated : non-aspirated). The adoption of these symbols for both phonetic and phonological representations has the drawback noted in Keating (1984: 287), e.g., that they

are used for a variety of physical events. Strictly speaking, [b d g] are supposed to be reserved for stops with voicing during their closures, and [p t k] for voiceless unaspirated stops. The symbols defined in this way can be

used in Polish, French, and many other languages quite straightforwardly. In contrast, their use in English and other languages is more complicated: [b d g] occur mainly in medial position, sometimes in initial position; [p t k] occur after [s], and sometimes medially. But /b d g/ are used as broad-transcription symbols for both [b d g] and voiceless [b̥ d̥ g̥], which occur mainly in initial position. The latter set has been called voiceless lenis, or voiceless lax, or even ‘voiced as in English’.

The reference to Indo-European plosive systems here is of more than tangential relevance to CyA. In his cursory survey of the role of aspiration in certain languages, Pétursson (1976: 169) remarks that

A l’exception du néerlandais les langues germaniques ont la caractéristique commune de présenter dans leur système consonantique le phénomène phonétique appelé aspiration. ... Le grec est aussi un exemple de langue européenne non-germanique possédant le trait phonétique de l’aspiration....

Distinctive aspiration is not ordinarily associated with the Arabic vernaculars. The display in CyA of distinctive aspiration and of voiceless, lax, dental stops (i.e., [t:] ⇐ /dd/)—features which this vernacular imbibed through contact with CyG—yields a striking phonetic parallel with some of the aforementioned Indo-European stop systems, especially Greek; it also underscores the necessity for a phonological analysis of CyA stops that goes well beyond a mere conventional classification of stops into phonetically voiced and voiceless ones, since, as the preceding discussion showed, no such phonological correlation can be postulated for CyA stops, there being no clear instances of spontaneous (i.e., non-conditioned) voicing of individual stop segments in this vernacular.

As Abercrombie (1967: 148–49) noted, though not with reference to CyA, “There can ... be many intermediate points ... at which voicing sets in: from ‘fully voiced’ to ‘voiceless fully aspirated’ is a continuum.” It would no doubt be very instructive to conduct an experimental study of CyA stop consonants in order to ensure clear and unassailable data on the exact function of voicing in this vernacular and on the extent of its correlation with other features (cf. Lisker and Abramson 1964, 1971; Chomsky & Halle 1968: 328ff.; Pétursson 1976; etc.). Such an investigation would almost certainly demonstrate the importance of laryngeal states other than vocal cord vibration to an analysis of the CyA stop series. In the absence of such a study I have resorted to a provisional phonological classification of CyA stops utilizing the tense : lax dichotomy familiar from traditional work on Indo-European languages, e.g., English.

15.7. Epilogue

All things considered, the special evolutionary path traversed by CyA, its continued interaction with Greek, and its present terminal state all suggest that it is probably unrealistic to approach the study of CyA phonology as if it were a completely integral and self-contained system in which, according to conventional wisdom, “tout se tient.” A more plausible representational model of CyA phonology must take cognizance of its “parasitical” relationship with the sound system of Greek and allow for the possibility that the organization of surface phonology in CyA is sometimes achieved at the expense of overall paradigmatic transparency. The aforementioned parallel case of the formal opacity of consonant length in Modern Hebrew is particularly relevant here, since this circumstance is a byproduct of discontinuity in language transmission and language contact. Another case of this kind, involving ambiguity in the interpretation of surface phonetics in a variety of Arabic, was, for instance, noted by Heath’s (1989) paper on the functional status of Moroccan affricates, in which the author concluded “that native speakers themselves are not entirely sure what is going on” (p. 133).

Thus, some systemically erratic traits in the voicing behavior of CyA stops noted in this chapter may ultimately also be symptomatic of this vernacular’s advancing obsolescence. A special study of the impact exerted on CyA phonology by its current sociolinguistic situation in Cyprus should perhaps take within its purview relevant comparative data from another terminal language surviving in a Greek socio-cultural context, i.e., Arvanitika (erstwhile Albanian) on mainland Greece (cf. Hamp 1989; Trudgill & Tzavaras 1977).

References

- Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press; Chicago: Aldine.
- Barth, Fredrik. 1969. “Introduction.” In *Ethnic Groups and Boundaries*, ed. Fredrik Barth, pp. 11–38. Boston: Little, Brown.
- Blanc, Haim. 1964. *Communal Dialects in Baghdad*. Cambridge: Harvard University Press.
- Borg, Alexander. 1985. *Cypriot Arabic*. Stuttgart: Steiner.
- Buhl, Frants. 1964 [1915]. *Wilhelm Gesenius’ Hebräisches und Aramäisches Handwörterbuch über das alte Testament*, 17th ed. Leipzig: Vogel.
- Cantineau, Jean, and Y. Helbaoui. 1953. *Manuel élémentaire d’arabe oriental*. Paris: Klincksieck.

- Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.
- Cohen, Marcel. 1912. *Le parler arabe des juifs d'Alger*. Paris: Champion.
- El-Hajjé, Hassan. 1954. *Le parler arabe de Tripoli (Liban)*. Paris: Klincksieck.
- Ferguson, Charles A. 1969. "The /g/ in Syrian Arabic: Filling a Gap in a Phonological Pattern." *Word* 25: 114–19.
- Fischer, Wolfdietrich. 1961. "Die Sprache der arabischen Sprachinsel in Uzbekistan." *Der Islam* 36: 232–63.
- Hagège, Claude. 1973. *Profil d'un parler arabe du Tchad*. Paris: Geuthner.
- Hamp, Eric. 1962. "The Interconnection of Sound Production, Perception, and Phonemic Typology." In *Proceedings of the Fourth International Congress of Phonetic Sciences held at the University of Helsinki, 4–9 September 1961* (Janua Linguarum Series Maior 10), ed. Antti Sovijärvi and Pentti Aalto, pp. 639–42. The Hague: Mouton.
- . 1989. "On Signs of Health and Death." In *Investigating Obsolescence: Studies in Language Contraction and Death*, ed. Nancy C. Dorian, pp. 197–210. Cambridge: Cambridge University Press.
- Heath, Jeffery. 1989. "Moroccan Affricates." In *Studia linguistica et orientalia memoriae Haim Blanc dedicata*, ed. Paul Wexler, Alexander Borg, and Sasson Somekh, pp. 133–35. Wiesbaden: Harrassowitz.
- Hopkins, Simon. 1990. Review of Borg 1985. *Zeitschrift für arabische Linguistik* 22: 88–93.
- Ingham, Bruce. 1994. "The Effect of Language Contact on the Arabic Dialect of Afghanistan." *Actas del congreso internacional sobre interferencias lingüísticas árabo-romances y paralelos extra-iberos celebradas en Madrid del 10 al 14 de diciembre de 1990*, ed. Jordi Aguadé, Federico Corriente, and Marina Marugán, pp. 105–17. Zaragoza: n.p.
- Jacoby, David. 1979. "Citoyens, sujets et protégés de Venise et de Gênes en Chypre du XIIIe au XVe siècle." In *Recherches sur la Méditerranée orientale du XIIIe au XVe siècle: Peuples, sociétés, économies* (Variorum Reprints 6), pp. 159–88. London.
- Jastrow, Otto. 1973. *Daragözü: eine arabische Mundart der Kozluk-Sason Gruppe (Südostanatolien): Grammatik und Texte*, Nuremberg: Carl.
- . 1977. "Gedanken zum zypriotischen Arabisch." *Zeitschrift der deutschen morgenländischen Gesellschaft* 127: 258–86.
- . 1978. *Die mesopotamisch-arabischen Qeltu-Dialekte*, vol. 1. Wiesbaden: Harrassowitz.

- Keating, Patricia A. 1984. "Phonetic and Phonological Representations of Stop Consonant Voicing." *Language* 60: 286–319.
- Kieffer, Ch. M. 1980. "L'arabe et les arabophones de Bactriane (Afghanistan)." *Welt des Islams* 20: 178–96.
- Levin, A. 1971. "The *imāla* in the Arabic Dialects." Ph.D. dissertation, Hebrew University, Jerusalem.
- Lisker, Leigh, and Arthur S. Abramson. 1964. "A Cross-language Study of Voicing in Initial Stops: Acoustical Measurements." *Word* 20: 384–422.
- . 1971. "Distinctive Features and Laryngeal Control." *Language* 47: 767–85.
- Lüdtke, Helmut. 1977. "Plurilinguismo e morfosintassi mista (comparazione del maltese con l'arabo cipriota)." *Bollettino dell'atlante linguistico mediterraneo* 18–19: 211–15.
- Newton, Brian. 1964. "An Arabic-Greek Dialect." In *Papers in Memory of George C. Pappageotes*, ed. Robert Austerlitz (Supplement to *Word* 20/3), pp. 43–52.
- . 1972a. *Cypriot Greek: Its Phonology and Inflections* (Janua Linguarum Series Practica 121). The Hague: Mouton.
- . 1972b. *The Generative Interpretation of Dialect* (Cambridge Studies in Linguistics 8). Cambridge: Cambridge University Press.
- Nöldeke, Theodor. 1904. Review of Stumme 1904. *Zeitschrift der deutschen morgenländischen Gesellschaft* 58: 903–20.
- Pétursson, Magnús. 1976. "Aspiration et activité glottale: Examen expérimental à partir de consonnes islandaises." *Phonetica* 33: 169–98.
- Röhrich, R. 1895. "Le pèlerinage du moine augustin Jacques de Vérone (1335)." *Revue de l'orient latin* 3: 178.
- Roth, Arlette. 1975. "Le verbe dans le parler de Kormakiti (Chypre)." *Epetêris* (Nicosia) 7: 21–117.
- Roth-Laly, Arlette. 1969. *Lexique des parlers arabes tchado-soudanais*. Paris: Centre national de la recherche scientifique.
- Schwyzer, Eduard. 1938. *Griechische Grammatik*, vol. 1. Munich: Beck.
- Selim, George Dimitri. 1967. "Some Contrasts between Classical Arabic and Egyptian Arabic." In *Linguistic Studies in Memory of Richard Slade Harrell*, ed. Don Graham Stuart, pp. 133–52. Washington, D.C.: Georgetown University Press.
- Singer, Hans-Rudolf. 1991. Review of Borg 1985. *Orientalistische Literaturzeitung* 86: 408–9.

- Sirat, A. S., and E. E. Knudsen. 1973. "Notes on the Arabic Dialect Spoken in the Balkh Region of Afghanistan." *Acta Orientalia* 35: 99–101.
- Storrs, Sir Ronald, and Brian Justin O'Brian, eds. 1930. *The Handbook of Cyprus*, 9th issue. London: Christophers.
- Stumme, Hans. 1904. *Maltesische Studien: Eine Sammlung prosaischer und poetischer Texte in maltesischer Sprache nebst Erläuterungen*. Leipzig: Hinrichs.
- Trudgill, Peter, and G. A. Tzavaras. 1977. "Why Albanian-Greeks Are Not Albanians: Language Shift in Attica and Biotia." In *Language, Ethnicity and Intergroup Relations*, ed. Howard Giles, pp. 171–84. New York: Academic Press.
- Tsiapera, Mária. 1969. *A Descriptive Analysis of Cypriot Maronite Arabic* (Janua Linguarum Series Practica 66). The Hague: Mouton.
- Ullmann, Manfred. 1970. *Die Medizin im Islam*. Leiden: Brill.

Maltese Phonology

Alexander Borg

Ben-Gurion University of the Negev

16.1. Introduction

The structural profile of contemporary Maltese represents the outcome of a diachronic process in which an erstwhile vernacular of spoken Arabic—brought to the Maltese Islands apparently in the late 9th century (A.D. 870 [?]; cf. Rossi 1936, Wettinger 1986, Barbato 1990)—evolved during most of the present millennium in virtually complete isolation from the mainstream Arabic vernaculars. Prolonged interaction with literary and dialectal varieties of Italian, the Mediterranean *Lingua Franca*—in the sense of Kahane and Tietze (1958)—and English, has acted as a strong catalyst of diachronic change accentuating the formal distinctiveness of Maltese (henceforth M) from native varieties of vernacular Arabic, particularly in the domains of lexicon, syntax, and phonology.

Though M is not infrequently referred to as an “Arabic dialect” in the linguistic literature, this designation would seem to be more appropriate as a genetic classification than as a synchronic one, since in its present form, M has structurally distanced itself far too profoundly from the norms of spoken Arabic to be regarded as anything other than a separate language. Two factors symptomatic of the highly alien character of the Maltese language community vis-à-vis the contemporary Arabic-speaking world deserve mention here:

(a) the Christian affiliation and European *Weltanschauung* of the Maltese, which renders them basically unreceptive to cultural and linguistic influences from the Arab countries. (Throughout its known history, the Maltese

AUTHOR’S NOTE: Abbreviations and symbols: /h/, in M words, stands for a phonemic segment that alternates freely between laryngeal, fricative, and velar points of articulation. The following abbreviations have been used: Ar., Arabic; Cal., Calabrian; Eng., English; It., Italian; Lat., Latin; M, Maltese; MLE, Mediterranean Lingua Franca; Mor., Moroccan; OA, Old Arabic; Sic., Sicilian; SM, Standard Maltese; Tk., Turkish; ⇐, derives synchronically from; ⇒, yields synchronically; { } enclose morphemic or allomorphic representations; | | enclose underlying representations. The letters of the M alphabet with notes relating to their most usual phonetic realizations are listed in § 16.8. I am indebted to Prof. A. Ambros (Vienna) for his comments on an earlier draft of this paper.

nation has stood distinctly outside the pale of such entities as *urūba* and *umma*.) The cultural and confessional aspects of the linguistic *Abstand* between M and Arabic are both strikingly reflected in the longstanding tradition of writing M by means of the Latin alphabet; cf. the use of the Hebrew alphabet in Judeo-Arabic.

(b) the absence of a literary Arabic tradition in Malta and the consequent non-existence of the so-called “diglossia” situation familiar from the Arab countries (cf. Ferguson 1959, but also El-Hassan 1977)—literary M being an independent local development deriving exclusively from colloquial M.

Thus whereas mainstream Arabic vernaculars form a symbiotic system with literary Arabic drawing on it both as a source of linguistic enrichment and as a cultural model, M lost its direct linguistic and cultural links with Arabic in the Middle Ages—presumably after the Norman invasion of 1090 and the subsequent expulsion of the Muslim community from the Maltese Islands in the 13th century (Wettinger 1986: 99)—and has since then looked to European languages (first Italian and later English) for cultural and lexical enrichment. Highly symptomatic of this profound and probably unbridgeable gap between M and the Arabic *Sprachraum* is the striking but rarely noted circumstance that notwithstanding its patently Arabic grammatical core, contemporary M is apparently unable, on both formal and cultural grounds, to draw on Arabic, e.g., as a source of lexical expansion.

A systematic attempt at assessing the extent of the formal drift of M from the Arabic dialect family and accounting for its apparent immunity against Arabic influences must take into account the aberrant evolutionary path that yielded the present M sound system. As in certain other residual varieties of dialectal Arabic spoken outside the Arab countries—e.g., in Cyprus (Borg 1985; Borg, this volume), and in Central Asia (Fischer 1961; Sirat and Knudsen 1973), all trace of velarization has been lost in the M consonant system, the OA emphatics having been fused with their plain counterparts. Far-reaching paradigmatic changes have also occurred among back consonants: reinterpretation of the OA velar and pharyngeal fricative pair ʕ and ɣ, chiefly as vocalic length; fusion of OA *x*, *ħ*, and *h* into M /h/, loss of *hamza*, etc. M has furthermore integrated several foreign consonant segments, primarily through contact with Italian: /p/, /v/, /c/, /dʒ/, and /ʃ/. As a result, the M sound system probably constitutes the most salient single factor accountable for the alien character and unintelligibility of the language to native speakers of Arabic. By the same token, the cumulative impact of these systemic departures from the phonological norms of spoken Arabic in M plau-

sibly represents the most important single correlate of the linguistic and cultural distance of its speaker community from the Arabic-speaking world, and of the *rapprochement* of M to the Romance *Sprachbund*.

Some idea of the estrangement of M from the Arabic *Sprachraum* in the perception of native Arabic speakers can be gleaned, for instance, from explicit statements by Arabic language reformers concerning the “decadent” state of M. Thus in a statement reflecting his concern for the purity of literary Arabic, Al-Maghribi (1947 [1908]: 43–44) characterizes “the language of Malta” as “neither Arabic nor foreign” (cited in Stetkevich 1970: 62). In the same spirit, the Syrian Anis Sallum wrote in 1922: “Is there an educated person who wants his language to be like that of the people of Malta?” (cited in Chejne 1969: 154). In a less formal vein, the Egyptian novelist Ibrahim al-Mazini (in *Šundu:q al-dunya*., 26–31) attempts to convey to the reader the foreignness of M by mimicking certain phonological traits in the language commonly associated with nonnative Arabic speech (e.g., confusion of Arabic /ħ/, /x/, and /h/).

The cultural distance of the Maltese nation from the Arabic-speaking world is clearly reciprocated in the self-perception of Maltese ethnic minorities abroad. Thus, whereas Maltese settlers in the West—England, Australia, and the U.S.—easily shift to the majority language, e.g., English (cf. Jeger 1963 cited in Dench 1975), M communities in North Africa and Egypt retained their ethno-religious distinctiveness (Nachtigal 1974 [1879]: 13–14; Vadala 1906, *passim*; Price 1954: 55f.) and their vernacular throughout their existence and did not simply shift to the local dialect of Arabic (cf. Saada 1986 on “Tunisian” Maltese, and Hull 1988 on “Egyptian” Maltese).

The phonology of M not only sets it apart from the mainstream Arabic vernaculars, but also constitutes the structural domain characterized by the greatest degree of internal dialectal differentiation, yielding a standard vernacular (henceforth SM)—spoken in the urban centers of the main island, Malta—and a number of nonstandard varieties traditionally associated with rural Malta and especially with the island of Gozo (for more detail, see Puech 1994: 16–22 and *passim*).

The consonantal inventory of M is virtually identical in all the dialects; fairly marked differences, however, obtain between SM and nonstandard M, as well as between the rural varieties themselves, in the domains of vocalic inventory and morphophonemics. Dialectal diversification in the rural vernaculars of M derives for the most part from differential historical treatment of the OA vowel system:

SM	Rabat (Gozo)	OA	
<i>omm</i>	<i>umm</i>	<i>umm</i>	‘mother’
<i>da:r</i>	<i>do:r</i>	<i>da:r</i>	‘house’
<i>bɪ:b</i>	<i>be:b</i>	<i>ba:b</i>	‘door’

Thus, whereas SM has evolved a system of five vowels, all nonstandard varieties of M have four (Puech 1994: 17f.). In the realm of vowel morphophonemics, certain non-standard varieties of M display complex rules of vowel harmony (Puech 1978) and/or a set of vocalic alternations conditioned by syllabic, lexical, and pausal boundaries. Observe, for instance, the prejunctural diphthongization of OA *[i:] and *[u:] in the M dialect of Rabat (island of Gozo), where the secondary diphthong in the closed syllable alternates with the monophthong in the open syllable:

Rabat (Gozo)	SM	
<i>zarbewn</i>	<i>zarbu:n</i>	‘shoes’
<i>zarbu:na</i>	<i>zarbu:na</i>	‘a shoe’
<i>treyd</i>	<i>tri:d</i>	‘you (sg.) want’
<i>tri:da</i>	<i>tri:da</i>	‘you (sg.) want her’

Certain conservative speakers from this dialect area also display pausally conditioned diphthongization of these OA long high vowels (Borg 1976)—a feature that is unknown in SM but is, significantly, well attested in certain Eastern varieties of vernacular Arabic, such as the dialect of Šhi:m (Lebanon; Fleisch 1974 [1962]: 203–20), and the North Palestinian dialect spoken by the Druze (Blanc 1953: 50).

The present sketch will focus primarily on the SM sound system as reflected in educated urban speech—e.g., among residents of Valletta, Sliema, etc.—and broadcast on the local radio and TV networks. Unless otherwise indicated, the immediate source of the SM linguistic corpus cited here is the present author’s own idiolect, and M dialectal material comes from the author’s fieldnotes.

16.2. The Consonants

16.2.1. *The consonant paradigm*

The SM sound system comprises the consonant segments in Table 16-1. Some consonants are restricted to a few but well integrated loanwords, e.g., /ʒ/ in /televíʒin/ ‘television’, /be:ʒ/ ‘beige’, /ru:ʒ/ ‘rouge’, /dz/ in /gadzdzetta/ ‘newspaper’, /landzi:t/ ‘bristle’, /medzdza/ ‘kind of basket’, etc.

Table 16-1. Maltese Consonants

p	t		k	'
b	d		g	
	f	s	š	h
	v	z	ž	
		c	č	
		dz	ǧ	
m	n			
	l			
	r			
w			y	

In addition to this consonantal paradigm, the M sound system has an “abstract phoneme” represented by the orthographical symbol *gh* (called [a:yn] in the M alphabet) and representing the reflex of the OA fricatives */ɣ/ and */ʕ/ (cf. § 16.2.4.12).

The schematic chart of M consonants displayed above characterizes the consonantal paradigm of SM. Certain nonstandard dialects show some minor differences; as noted in Puech (1994: 17), conservative speakers from the Gozitan villages of San Lawrenz, Għarb [a:rp], and Zebbuġ [zebbú:ǧ] still retain residual traces of the voiced velar fricative [ɣ] continuing the same phoneme in Old Arabic. Archaic consonantal traits have also been retained by speakers on the island of Malta itself. Thus, in contrast with the glottal reflex of *qāf* in SM, conservative speakers from the towns of Cospicua and Senglea [séngleya] retain the uvular stop [q]—also noted for Valletta in the early years of the century in Stumme 1904 (passim). The dialect of Xewkija [šewkiyya] in Gozo, on the other hand, systematically fuses OA /q/ and /k/ into /k/ (for further detail, see Puech 1994: 31). Dialectal variation also extends to M reflexes of OA /r/; instead of the apical trill [r] that is the normal SM reflex of OA *r, some nonstandard dialects of M show the alveolar tap [ɾ] or the alveolar approximant [ɹ] (cf. Puech 1994: 17).

16.2.2. *Phonological neutralization in consonants*

16.2.2.1. *Voicing*

Voiceless stops in M are generally realized without a significant degree of aspiration when released, and voiced stops tend to be uttered with full

phonation outside automatically devoicing phonetic environments. Some speakers display optional secondary voicing of historically voiceless stops in resonant environment in certain Romance loans. The integral function played by the voicing opposition in the M obstruent series as a result of its integration of the new obstruents *p*, *v*, *č*, *g*, *c*, and *ž* (acquired both through internal change and via nativized loans) is also reflected to varying degrees in many mainstream Arabic vernaculars that have extended the paradigmatic range of the Old Arabic voicing contrast.

M obstruents are subject to a regressive voicing assimilation rule systematically converting all and every dyadic or triadic set of adjacent obstruent segments that are nonuniform in voicing into completely voiced or voiceless sequences. The voicing feature in Maltese is therefore non-distinctive in all but the last segment in a cluster. Automatic gain or loss of the voicing feature in obstruents has given rise to conditioned alternation between voiced and voiceless segments that can otherwise occur in positions of functional contrast, e.g., the voiced and voiceless bilabial stops /p/ and /b/:

[kítep]	‘he wrote’	⇐ /kiteb/
[nígdbu]	‘we write’	⇐ /niktbu/
/posta/	‘mail’	< It. <i>posta</i>
/bosta/	‘many’	< Ar. <i>bašta</i>

The voicing opposition in M is also neutralized in prejunctural position (as in [kítep] ‘he wrote’), where voicing contrasts are normally restricted to underlying representations. Word-final devoiced obstruents in M are phonetically indistinguishable from their underlying voiceless correlates, but will here be transcribed as phonologically voiced so as to distinguish them from historically voiced obstruents in which the voicing feature has been permanently lost: /art/ ‘land’ < Ar. *arḏ*, /niket/ ‘sorrow’ < Ar. *nakd*, etc. The present distribution of voicing in neutralizing environments is not accurately reflected by contemporary orthographical norms since the Maltese writing system tends to be etymological in this respect; thus it often indicates obstruent voicing in contexts where permanent neutralization has rendered it irretrievable to the contemporary speaker: <sibt> [sɪpt] ‘Saturday’ < Ar. *sabt*, <abt> [a:pt] ‘armpit’ < Ar. *ibt*, etc. The tendency towards hypercorrect voicing is also reflected in orthographical variants, such as <berghud/berghut> ‘flea’ < Ar. **baryu*:θ.

In effect, voicing neutralization in M has tended to obscure native speaker perceptions of the underlying and historical distribution of voicing in obstruents. This circumstance has occasioned re-etymologization of several M

root morphemes, entailing, for instance, permanent gain or loss of the voicing feature in certain obstruents via back formation. Observe the following examples, where conditioned voicing of the M reflexes of OA **k* and **š* in clusters has yielded permanent M /g/ and /z/: /gideb/ 'he lied' < Ar. **kaðab*, /zebah/ ~ /zeba/ 'he painted' < Ar. **šaba*y. Permanent devoicing of historically voiced segments is also common: /tefa/ 'he threw' < Ar. **dafa* 'to push', /hafer/ 'he forgave' < Ar. **yafar*, etc.

The propensity for this kind of back-formation in M has occasioned numerous instances of hypercorrect shifts, as in the case of /d/ in M /disa/ 'nine' (< Ar. *tis'a*) back-formed from /sata:š/ 'nineteen' popularly perceived as ldsata:šl.

16.2.3. Length

M has retained the OA length contrast in both vowels and consonants. Consonantal length is optimally retained intervocalically after stressed vowels:

<i>ferah</i>	'he rejoiced'
<i>ferrah</i>	'he made happy'
<i>safar</i>	'travel'
<i>saffar</i>	'he whistled'

Both in lineally inherited lexicon and in nativized loans, M displays a certain degree of interdependence between the length of a vowel and that of a following word-final consonant. Thus, in monosyllabic words, long M vowels inherited from OA occur optimally before single consonants:

<i>ġa:r</i>	'neighbor'
<i>ġarr</i>	'he dragged'
<i>bɪ:l</i>	'he urinated'
<i>bill</i>	'dip!'
<i>vo:t</i> (< It. <i>voto</i>)	'vote'
<i>bott</i> (< Sic. <i>botti</i>)	'tin can'

Rare contraventions of this canonic norm occur in the few cases where M has evolved secondary vowel length before a word-final CC sequence:

<i>e:mm</i>	'there' < * <i>hemm</i> < Mor.Ar. <i>hamm</i> (Fischer 1959: 127 n. 2)
<i>e:kk</i>	'thus' < * <i>hekk</i>

M tends to neutralize consonantal length in prejunctural position:

[fomm-i]	'my mouth'
[fom]	'mouth'

Stressed phonetic -VC# word-final sequences in M derive uniquely from underlying /-VCC#/, since singly closed prejunctural syllables displaying short

vowels are rarely if ever stressed. Thus, as in many Arabic dialects, foreign loans displaying stressed word-final -VC# in the lending language tend to undergo automatic gemination of their final consonant in M underlying representations: /klabb-iyɪ:t/ ‘club-s’, /čekk-iyɪ:t/ ‘check-s’, etc. This constraint on the syllabic structure of monosyllables in M occasionally yields free variation between the canonic forms CVCC and CV:C, e.g., in the loanword /ramm/ ~ /ra:m/ ‘copper’ < Italian *rame*.

Many Arabic dialects tend to neutralize the length contrast in consonant clusters. Schabert (1976: 26) notwithstanding, however, many M speakers, including the present writer, often retain such contrasts, possibly under the influence of written norms:

<i>nizlu</i>	‘they descended’
<i>nizzlu</i>	‘they brought down’
<i>čahdet</i>	‘she denied’
<i>čahhdet</i>	‘she deprived’

Underlying and historical contrasts between long and ultra-long consonants are, however, always automatically neutralized:

<i>šellel</i>	‘he basted’
<i>šellet</i>	‘she basted’ $\Leftarrow$ šelllet $\Leftarrow$ šellel + et
<i>hazzez</i>	‘he scribbled’
<i>hazzet</i>	‘she scribbled’ $\Leftarrow$ hazzzet $\Leftarrow$ hazzez + et

16.2.4. Sound Classes

The present section illustrates the contrastive roles of M consonants and outlines pertinent phonetic and distributional features. Diachronic aspects of the M sound system are briefly touched upon whenever this can throw light on present distributional patterns or special morphophonemic processes.

16.2.4.1. Labials

The M labial series comprises the bilabial stops /p/ and /b/, the nasal bilabial /m/, and the labiodental fricatives /f/ and /v/:

SM	<i>post</i>	‘place’
	<i>bosta</i>	‘many’
	<i>fost</i>	‘among’
	<i>mošt</i>	‘comb’
	<i>pošt</i> (< Tk. <i>pušt</i>)	‘scoundrel’

<i>rkoppa</i> (< Ar. <i>rukba</i>)	'knee'
<i>kobba</i>	'ball of wool'
<i>'offa</i>	'wicker basket'
<i>komma</i>	'sleeve'

The extent to which M has integrated the voiceless bilabial stop /p/ in its sound system is probably without a close parallel among the contemporary Arabic dialects, even those where the voiceless bilabial stop is a full denizen (e.g., Baghdadi Arabic, Blanc 1964: 18). M has, for instance, not only integrated foreign /p/ but has proliferated it in new lexical creations, especially in words of an onomatopoeic nature:

SM	<i>čappas</i>	'he tainted'
	<i>tappan</i>	'he made dull (glass, mirror)'
	<i>perper</i>	'it fluttered in the wind (flag)'
	<i>pačpač</i>	'he chattered'
	<i>pa'pa'</i>	'he sounded the horn (of car)'
	<i>pespes</i>	'he whispered'

16.2.4.2. Labiodental fricatives

The labiodental fricative /v/ is understandably rare in lineally inherited words; note, however, its occurrence in the high-frequency SM term /i:va/ 'yes' < **aywa*, which yields a functional contrast with /f/ in /ği:fa/ 'fool'. A variant [i:wa] (generally perceived as nonstandard) also exists. The diachronic shift **w* > *v* here presumably reflects the co-articulatory incompatibility of lip-spreading concomitant with [i:] and lip-rounding accompanying [w]. In fact, the sequence [iw] is altogether rare in M (cf. § 16.3.3).

The contrast between M /f/ and /v/ is well established by reason of numerous Italian loans displaying the voiced labiodental fricative in initial and medial positions:

<i>fi:ni</i>	'fine (pl.)'	< It. <i>fini</i>
<i>vi:ni</i>	'veins'	< Sic. <i>vini</i>
<i>fe:ra</i>	'he injured'	< It. <i>ferire</i>
<i>ve:ra</i>	'true (f.)'	< It. <i>vera</i>
<i>na:fu</i>	'we know'	< Ar. * <i>na'rafu</i> :
<i>tra:vu</i>	'beam'	< Sic. <i>travu</i>

In M the voiced labiodental fricative /v/ does not enjoy the distributional freedom of most other M consonants and, for instance, rarely occurs geminated outside certain morphological contexts, e.g., verbal inflection: /vvo:ta, yivvo:ta/ 'to vote'.

16.2.4.3. *Dentals*

M /t/ (< OA *t*, *θ*, and *t̥*) and /d/ (< OA *d*, *ð*, *d̥*, and *ð̥*) are usually realized as plain dental stops, whereas the segments /s/ (< OA *s* and *ʃ*), /z/, /n/, /l/, and /r/ all are articulated with a more retracted tongue position, i.e., against the alveolar ridge. The following word pairs illustrate functional contrasts within this sound class; it should, however, be borne in mind that M neutralizes the voicing opposition word-finally in obstruents:

<i>t : d</i>	<i>temm</i>	‘he finished’	<i>demmm</i>	‘blood’
	<i>’atel</i>	‘he killed’	<i>’adef</i>	‘he rowed’
	<i>wiret</i>	‘he inherited’	<i>bired</i>	‘it (masc.) cooled’
<i>s : z</i>	<i>smɪ:n</i>	‘fat (pl.)’	<i>zmɪ:n</i>	‘time’
	<i>nisel</i>	‘origin’	<i>nizel</i>	‘he descended’
	<i>biss</i>	‘only’	<i>wizz</i>	‘geese’

16.2.4.4. *Affricates*

The M alveolar series also includes the voiced and voiceless affricates /c/ and /dz/ respectively:

<i>c : z</i>	<i>copp</i>	‘lame (masc.)’	< Sic. <i>zoppu</i>
	<i>dzo:na</i>	‘zone’	< It. <i>zona</i>
	<i>pecca</i>	‘roll of cloth’	< Sic. <i>pezza</i>
	<i>meddzza</i>	‘kind of basket’	< It. <i>mezza</i>
	<i>pasticc</i>	‘cheesecake’	< Sic. <i>pastizzu</i>

Both segments tend to occur primarily in words of Italian provenance and usually retain their phonological length component in non-initial position. Observe, for instance, how the segment /c/ in fully integrated italianisms functions as a geminated radical when the root undergoes internal inflection: /bocca/ ‘electric bulb’ (pl. /bococ/) < It. *bozza* ‘protuberance’. The voiceless affricate /c/ also shows up in M derivatives of Arabic *kuds* (pl. *akda:s*) ‘heap’, which yields M /gocc/ (pl. /gcu:c/) ‘heap’, and the verbal forms /geccec, ygecc-ec/ ‘to bundle together’.

16.2.4.5. *Resonants*

Aside from a few noteworthy cases of permutation and metathesis discussed in § 16.2.4.7 below, the OA resonants /l/, /m/, /n/, and /r/ have been largely retained without change in lineally inherited lexicon, where they contrast functionally in word-initial, medial, and final positions:

l : m : n : r :

<i>laha</i> 'he reached'	<i>mela</i> 'he filled'	<i>dahal</i> 'he entered'
<i>mahat</i> 'he blew (nose)'	<i>rema</i> 'he threw away'	<i>šaham</i> 'grease'
<i>naha</i> 'it brayed'	<i>fena</i> 'he weakened'	<i>tahan</i> 'he ground'
<i>rahal</i> 'village'	<i>mera</i> 'mirror'	<i>nahar</i> 'he snored'

The nasal continuants /n/ and /m/ do not contrast before stops but always agree with the latter in point of articulation:

<i>yonfoħ</i>	[yomfoħ]	'he blows'
<i>banda</i>	[banda]	'side'
<i>bank</i>	[baŋk]	'bank'

The phonological neutralization of the labial/dental contrast in nasals before stop segments yields the nasal archiphoneme *N*; this abstract representation is borne out by the tendency of M speakers to extrapolate underlying /n/ from surface [mʃ] and [mp] clusters: M [ˈnɪ:pən] 'bells', sg. [ˈampɪ:na] < It. *campana*.

The M alveolar nasal /n/ also assimilates to stem-initial resonants in the imperfect paradigm of verbs: /mmu:r/ 'I go' ⇐ /nmu:r/, /rri:d/ 'I want' ⇐ /nri:d/, etc. In the dual or pseudo-dual ending {-eyn} the final /n/ is liable to alternate freely with zero, usually with concomitant secondary doubling of the preceding glide or, alternatively, lengthening of the steady state portion of the vowel nucleus: /tneyn ~ tneyy ~ tne:y/. The elision of /n/ here and in reflexes of Ar. *-ayn is presumably conditioned by the synchronic morphological boundary between /y/ and /n/ in the pseudo-dual: /idey-h/ 'his hands', /sa'ay-h/ 'his legs', etc. Note also, in this connection, the historical accretion of word-final /n/ in M /armayn/ 'too late' < It. *ormai*, and its proliferation as a secondary syllable coda after word-final long stressed vowels, where it presumably implements a strategy for integrating foreign words, e.g., by facilitating morphological inflection of nouns: /sufá:n/ (pl. /sufan+iyɪ:t/) 'sofa' < It. *sofà*, /blu:n-a/ 'indigo' < Eng. *blue*, /skru:n/ (pl. /skreyyen/) 'propeller' < Eng. *screw*, /kafé/ 'coffeehouse' (pl. /kafen-iyɪ:t/).

16.2.4.6. The Lateral Resonant

The lateral liquid /l/ in the inherited lexicon is generally invariable at the purely segmental level but is subject to grammatically conditioned alternation at morphemic level, e.g., as a definite article and as object marker. The definite article {(i)l-} generally assimilates to following coronal segments (*t*, *d*, *s*, *z*, *n*, *r*): /it-ti:n/ 'the figs', /iš-šemš/, 'the sun', /ir-ra:s/ 'the head', /in-na:r/ 'the fire', /id-da:r/ 'the house', etc.; interestingly, the assimilation rule is

asymmetrically implemented before the voiced and voiceless alveopalatal affricates /ǧ/ and /č/, respectively: *il-ǧisem* ‘the body,’ but *ič-čarru:ta* ‘the rag’ (~ Ar. *šarṭu:ta*). Non-assimilation of the lateral before /ǧ/ is the norm in Classical Arabic and certain Arabic colloquials.

In some Eastern Arabic dialects the definite article is subject to optional assimilation before [ǧ]: rural Palestinian [il- ~ iǧ-ǧa:r] (Bauer 1926: 5); this suggests that the voiced alveopalatal affricate is ambiguously categorized in the consonant systems of these vernaculars. Assimilation of /l-/ exclusively before the voiceless alveopalatal affricate in M may well hark back to a historical stage when this affricate was perceived by native M speakers as the cluster [tʃ]. (For further discussion of this point, see Comrie 1980.)

Like many mainstream Arabic dialects, M displays the optional assimilation of /l/ to a following /n/ exclusively at the morpheme boundary between a verb stem and the pronominal suffix {-na:} ‘we, us’: /kilna ~ kinna/ ‘we ate’, /hadilna ~ hadinna/ ‘he took from us’, etc. Another morphologically conditioned realization of /l/ occurs in the (direct and indirect) object marker {lil-}, which is prone to automatic gemination before an initial vowel: /lill-ommi/ ‘to my mother’ ⇐ lil-lómmil (but /lil-missɪ:ri/ ‘to my father’), thus avoiding an open unstressed syllable.

16.2.4.7. *Interaction between Lateral and Trill*

In words of Italian origin there is an unsystematic though widespread synchronic trend (especially in certain very commonly used words) towards re-shuffling of liquids such that the lateral occurs optimally in syllable-initial and the apical trill in syllable-final position. Historically, this has yielded permanent syllabic shifts such as the following:

SM		Italian
<i>arblu</i>	‘mast’	< <i>albero</i>
<i>porvli</i>	‘gunpowder’	< <i>polvere</i>
<i>girgo:r</i>	(name)	< <i>Gregorio</i>
<i>perli:na</i>	‘sugar-coated almond’	< <i>pralina</i>
<i>garigo:r</i>	‘spiral staircase’	< Sic. <i>caragolu</i>
<i>gabiryo:la</i>	‘somersault’	< <i>capriola</i>
<i>gamblu</i>	‘crayfish’	< Cal. <i>gambaru</i>
<i>splengu:n</i>	‘brooch’	< Sic. <i>spinguluni</i>

Though rare in the Arabic lexical component, this type of syllabic restructuring may account for the root-final changes in the native words

M		Arabic
<i>tamal</i>	'dates'	<i>tamar</i>
<i>fɪ:tel</i>	'lukewarm'	<i>fɑ:tir</i>

presumably via back-formation from the feminine forms (*tamla*, *fɪ:tla*) showing the resonant segment in syllable-initial position.

In M lexicon of Romance provenance the same distributional preference can also be observed in the historical diffusion of *secondary* liquids promoting closed syllabicity in unstressed positions, especially, though not exclusively, in the nautical lexicon: /skorfi:na/ 'nut (mech.)' < Sic. *scufina*, /arcnell/ (kind of fish) < M **arsnell* ~ Sic. *asineddu*. Note also the proliferation of word-final /r/ from /l/, other resonants, or zero: /arta:l/ 'altar' < It *altare*, /pinnu:r/ 'weathercock' < It. *pennone* 'flagpole', /gʊvintu:r/ 'young men' < Sic. *guivintu*, etc.

Synchronically, this process is most noticeable in the speech of the uneducated, where the restructured forms implement the unmarked syllabic option: [pitlóryu] 'paraffin', [sturmént], 'musical instrument' [maskla] 'mask', vs. SM [pitrólyu] [strumént], [maskra], ultimately < It. *petrolio*, *strumento*, and *maschera*. The popular perception of secondary /r/ (< another resonant or zero) at syllable codas as a hallmark of uneducated M speech is sometimes consciously exploited in the dialogue sections of M literary works portraying life in rural Malta: *ipperparot* 'prepared' < It. *preparato*; *porpaganda* 'propaganda' (Mamo 1930–31: 24, 4). Certain nonstandard M speakers use the form /telefórnr/ 'telephone', where insertion of the apical trill creates a secondary CC cluster regularizing word-final stress in this lexical item (for the M stress rule, see § 16.5 below).

In standard M speech, this process yields a certain amount of free variation between /l/ and /r/ at syllable codas: /gʊrf/ ~ /gʊlf/ 'giant' < (?) Ar. *gurf* 'cliff', /galbu/ ~ /garbu/ 'good manners' < It. *garbo*.

16.2.4.8. Alveopalatals

The M alveopalatal series comprises principally the voiceless fricatives /s/, and the voiced and voiceless affricates /gʃ/ and /č/, respectively. The following pairs illustrate the relevant contrasts:

<i>s</i> : <i>š</i>	
<i>sellef</i> 'he lent'	<i>šellef</i> 'he blunted (blade)'
<i>nesa</i> 'he forgot'	<i>meša</i> 'he walked'
<i>flu:s</i> 'money'	<i>tru:š</i> 'deaf (masc.)'

š: č: ġ

šaham 'grease'

čahad 'he denied'

ġahan 'Juha'

kašša 'box'

kačča 'hunt (n.)'

gaġġa 'cage'

baraš 'he scratched'

ar'ač 'he scorched'

taraġ 'steps'

Though initially M /č/ may well have been internally generated via historically bi-phonemic *tš or devoiced *ġ in the inherited lexicon, the integration into M of a large number of Italian loans displaying the alveopalatal affricate in different positions had the effect of regularizing its distribution:

čavetta

'key'

kečča (< It. *cacciare*)

'he expelled'

a'očč (< Sic. *cacocciula*)

'artichokes'

M nonetheless retains the formal distinction between the sequence /tš/ and /č/, since the former usually implies a morphological boundary between stop and fricative:

ma-ra:t-š

'she did not see'

Thus, for instance, whereas for the purposes of stress assignment, a word-final /tš/ represents a consonant cluster and draws stress on a preceding short vowel, non-geminated /č/ does not:

ma-čartít-s

'she did not rip'

ármeč

'he crunched'

The voiced alveopalatal affricate /ġ/ has the optional allophone [ž] before stops, e.g., [ğdi:d] ~ [ždi:d] 'new (masc.sg.)'. Loss of the occlusive onset probably represents a process of cluster simplification: [džd] > [žd]. It is presumably the synchronic alternation between [ğ] and [ž] in this phonetic context that has had the effect of obscuring the historical phonemic boundary between reflexes of /ğ/ and /š/: [ğbi:n] ~ [žbi:n] 'chum, fellow', [ğbayt] ~ [žbayt] 'I am sated', ultimately < Ar. *šbi:n 'godfather' and *šabi't, respectively.

16.2.4.9. Velar Stops

The opposition of the voiceless velar stop /k/ and of its voiced equivalent /g/ in M was generated internally within the Arabic-based lexicon via voicing assimilation, both progressive and regressive:

k: g

kerrah 'he made ugly'

(~ Ar. *kariha* 'he hated')

gerrem 'he gnawed (bone)'

(cf. M *nqarram 'I gnaw'

~ Ar. *qarama* 'he gnawed')

tnikker 'he loitered'

niggez 'he pricked'

(cf. M *nikkzu ~ Ar. *nakaz*)

and has been reinforced by numerous Italian words displaying /g/ in most environments. The contrast between /k/ and /g/ is partially neutralized in contact with certain resonants, e.g., liquids: *sa:gru* ~ *sa:kru* ‘holy’ < Sic. *sacru*, *klassi* ~ *glassi* ‘class’ < It. *classe*, *kazi:n* ~ *gazi:n* ‘club’ (cf. *il-gazi:n*, ultimately from It. *casino*), etc.

16.2.4.10. M /g/ vs. /ǧ/

Occasional traces in M of a shift /ǧ/ > /g/ in the context of /z/ (e.g., /gzi:ra/ ‘island’ < Ar. *ǧazi:ra*)—a well known feature in certain Arabic dialects—suggest that phonologization of the *ǧ*: *g* contrast in Maltese was, at least in part, an inherited trait. Maltese appears to have itself expanded the lexical scope of this contrast, both via internal change (e.g., OA **q* and **k* > M *g*):

<i>g</i> : <i>ǧ</i>	M		Arabic
	<i>garr</i>	‘he complained’	< <i>qarr</i>
	<i>ǧarr</i>	‘he carried’	
	<i>niggez</i>	‘he pricked’	< <i>nakaz</i>
	<i>niǧǧes</i>	‘he fouled’	

and through subsequent heavy borrowing from Mediterranean languages and English: /ganč/ ‘hook’, /bagoll/ ‘suitcase’, /sso:gra/ ‘he risked’ (< MLF; cf. Ar. /so:gar/ ‘he insured’), /gendu:s/ ‘bull’ (< Berber), /gara:š/ ‘garage’ (< Eng.), etc.

16.2.4.11. The M fricative /h/

As already noted, M has lost the distinctions between the OA voiceless pharyngeal, velar, and laryngeal fricatives **ħ*, **χ*, and **h*, respectively. The unconditional merger of **ħ* and **χ* had probably already occurred in Medieval M; note, for instance, the use of the digraph *ch* for the reflex of both OA sounds in the late 15th century poem, *Peter Caxaro’s Cantilena* (Wettinger 1968): <bachar> ‘sea’ < *baħr*, <morchi> ‘loose’ < **marxī*. The contemporary orthography renders this consonant by <h>; in the present study, it is indicated by plain /h/.

Orthography	Phonemic representation		Arabic
<i>ħajt</i>	/hayt/	‘wall’	< * <i>ħa:yiṭ</i>
<i>ħajt</i>	/hayt/	‘thread’	< <i>xayṭ</i>
<i>baħar</i>	/bahar/	‘sea’	< <i>baħr</i>
<i>daħal</i>	/dahal/	‘he entered’	< <i>daxal</i>
<i>riħ</i>	/ri:h/	‘wind’	< <i>ri:h</i>
<i>xih</i>	/ši:h/	‘old man’	< <i>šayx</i>

Many M speakers tend to articulate /h/ as a pharyngeal fricative [ħ] and it is usually transcribed as such (ħ) in descriptive work, e.g., Stumme (1904, passim), Schabert (1980: 287f.), and Puech (1994, passim), but other pronunciations, for instance as a voiceless velar fricative [x], or as a voiceless laryngeal fricative [h], also exist. These phonetic variants of /h/ do not appear to correlate with speakers' sociolectal backgrounds or with dialect boundaries. In the present author's usage, non-laryngeal pronunciations of /h/ are the norm in the context of gemination: [kahħal] 'he plastered'.

Consonantal reflexes of the OA laryngeal fricative *ħ in M are phonetically identical with those of OA *h and *x. The laryngeal fricative has nonetheless undergone a distinctive evolutionary process in M entailing large-scale elision and/or fusion with historically adjacent short vowels in certain contexts. This has yielded secondary vocalic length, as in

M		Arabic
<i>e:dded</i>	'he threatened'	<i>haddad</i>
<i>e:na</i>	'joy'	<i>hana:ʔ</i>

or morphophonemic alternation between M /h/ and zero:

Orthography	Phonetic realization	
<i>iblah</i>	[íblah]	'foolish (masc.)'
<i>belha</i>	[béla]	'foolish (fem.)'
<i>ikrah</i>	[íkrah]	'ugly (masc.)'
<i>kerha</i>	[kéra]	'ugly (fem.)'

The M orthography tends to be etymological with respect to OA /h/ and often shows <h> where its consonantal value has been entirely lost:

Orthography	Phonetic realization	
<i>rahan</i>	[ra:n]	'token'
<i>huwa</i>	[úwwa]	'he'
<i>hija</i>	[íyya]	'she'
<i>bhejjem</i>	[béyyem]	'animals' (< *bħa:yim)

In a few residual cases, however, OA /h/ has yielded a phonemically stable consonant /h/: /nahaʔ, yinhaʔ/ 'to bray', /nebbah, inebbah/ 'to enlighten'. Traditionally the M orthography also tended to retain the etymological distinction between the reflex of OA /ħ/ and /x/—which it represents by <ħ>—and that of stable /h/, written <h>:

Orthography	Phonetic realization	
<i>ħaffer</i>	[ħáffer]	'he dug'
<i>harab</i>	[ħarab]	'he escaped'

However, the most recent dictionary of M, Aquilina (1987–90), follows the current trend in literary M and writes all instances of the stable radical /h/ as <h> irrespective of etymological source.

16.2.4.12. *Maltese ghajn*

The term *ghajn* refers to the digraph <gh> used in the M writing system for representing the reflexes of both OA /ʕ/ and /ɣ/. In his table of M consonants, Cohen (1970 [1966]: 127) designates this segment as a pharyngeal but later notes that the orthographical digraph ‘au moins à La Valette et dans les villes, ne semble pas correspondre toujours à un trait acoustique audible’ (p. 129).

In fact, the most common phonological reflex of these OA consonants in standard and non-standard varieties of M is vocalic length:

Orthography	Phonetic realization	OA
<i>ghamel</i>	[a:mel]	<i>ʕamal</i>
<i>ghalaq</i>	[a:laʔ]	<i>ɣalaq</i>
<i>soghla</i>	[so:la]	<i>suʕla</i>
<i>ragħwa</i>	[ra:wa]	<i>raywa</i>

A few conservative rural communities (referred to in § 16.2.1) have nonetheless retained in a residual fashion the OA voiced velar fricative [ɣ]. In SM and most other varieties of M, consonantal reflexes of *ghajn* are restricted to certain morphophonemic contexts, e.g., before the suffixed pronouns {-ha:} and {-hom} in verb inflection., where historical or underlying *ghajn* is realized as /h/:

M	<i>sema</i>	‘he heard’	< * <i>sama</i> ‘ < OA <i>sami</i> ‘
	<i>semahha</i>	‘he heard her’	< * <i>samaʕha</i> :

As can be inferred from the second syllable of M [séma], in positions where M neutralizes vowel length (e.g., prejuncturally), no surface reflex of these OA sounds occurs. Synchronic alternation of this kind is not restricted to M verbs but can also occur in nominal inflection; observe the morphophonemic alternation /h/ ~ /ø/ ~ /y/ in the following nouns:

M		OA
[déma], [dmu:h]	‘tear(s)’	<i>damʕa</i> , <i>dumu:</i> ‘
[ʔi:h], [ʔiya:n]	‘bottom(s)’	<i>qa:</i> ‘, <i>qi:a:n</i>

Incidentally, orthographical *gh* in M does not always correlate synchronically with a specific phonological element in underlying phonology; thus it sometimes shows up as a purely otiose etymological residue in a number of M words (especially nouns) where its phonetic reflexes, primary or

secondary, have been entirely lost: <naghal> [na:l] ‘horseshoe’ < **naʿal* < OA *naʿl*, *xoghol* [ʃo:l] ‘work’ < **ʃoyol* < OA *ʃuyl* (both M words are monosyllabic). In some cases, orthographical *ghajn* in M lacks an etymological basis altogether and is clearly the product of analogical change: *laqha* [laʿa] ‘meeting’ (< Ar. *liqaʾ*), *jilqagħhom* [yilʾāħhom] ‘he receives them’, *baqghu* [báʿaw] ‘they remained’, etc.

The resort to a digraph *gh* in the contemporary M orthography is an elegant practical solution for capturing the complex synchronic morphophonemics arising out of the phonetic reinterpretation of OA /ʕ/ and /ɣ/ in the evolution of M. Thus basing himself principally on M verb morphophonemics, Brame (1972) concluded that setting up an abstract pharyngeal segment *ghajn* in the M consonant paradigm facilitates the statement of formal regularities relating, for instance, to M word stress, distribution of stem vowels, etc., in this form class.

Phonological evidence in support of the synchronic reality of an abstract segment in M includes the noteworthy fact that for many M speakers secondary long vowels harking back to *‘+ V (or vice versa) are phonologically marked and, unlike primary long vowels, are not subject to automatic reduction (in conformity with the length neutralization rule stated in § 16.3.2) when they undergo stress shift:

<i>á:mmar</i>	‘he dwelt’ (< Ar. <i>ʿammar</i>)
<i>a:mmárt</i>	‘I dwelt’

It should, however, be noted that the generative interpretation of abstract *ghajn* may ultimately prove more faithful to historical fact than to the synchrony of M. At all events, the notion that M speakers perceive an underlying ‘pharyngeal segment’ *ghajn* finds little support in written usage, since the correct assignment of this digraph in written M remains a notorious source of error even among highly literate speakers. For an alternative analysis of M *ghajn*, see Comrie (1986).

16.2.4.13. *The glottal stop*

Like many North African Arabic vernaculars, M no longer retains the automatic glottal catch preceding word-initial vowels: M /isem/ ‘name’, /omm/ ‘mother’ (not /ʾisem/, /ʾomm/), though occasional secondary vocalic length in this context may well hark back to OA *hamza*: /a:bt/ ‘underarm’ < ʾibt/, /ewlɪ:mes/ ‘the day before yesterday’ < ʾawwal ʾams.

M /ʔ/ in standard speech represents a radical consonant usually continuing OA *qāf*, though certain nonstandard speakers (e.g., from Victoria and Xew-

kija, Gozo) fuse OA /q/ and /k/ into /k/. It is probably hypercorrect interference from these vernaculars that has yielded the SM variants /aː/, /kaːk/ ‘round cakes’ < Ar. *kaʔ*, SM /serduː/, /serduːk/ ‘rooster’ < Maghrebi Ar. *sar-duːk*, etc.

The functional contrasts between SM /ʔ/ and other backed consonants can be exemplified as follows:

ʔ: k	ʔattar	‘it dripped’	kattar	‘he increased’
	baʔra	‘cow’	sakra	‘drunkenness’
	toroʔ	‘streets’	torok	‘Turks’
ʔ: h	ʔassar	‘he shortened’	hassar	‘he erased’
	baʔar	‘cows’	bahar	‘sea’
	baʔ	‘bedbugs’	bahh	‘nothing’

16.2.5. Phonaesthemes

Some M consonants, principally /č/, /c/, /p/, and /ʔ/, sometimes assume a special expressive value, in the sense of Firth (1964: 184). The rich inventory of onomatopoeic lexical creations in M, particularly in the quadriliteral verb class, is of some interest in a phonological description since the spontaneous origin of such terms attests to the articulatory naturalness and phonological integration of these segments in the M sound system:

čarrat	‘he ripped (cloth)’ (cf. Ar. <i>šaraʔ</i>)
čekček	‘he rattled’
čapčap	‘he clapped’
čafčaf	‘he splashed water’
ʔarmeč	‘he crunched (food)’
mečlaʔ	‘he ate noisily’
cekcek	‘he tsked’
ci:ci	‘poultry (nursery)’
paʔpaʔ	‘he honked (the horn of a car)’
pačpač	‘he chattered’
perper	‘it fluttered in the wind (flag)’

M /č/ also has other expressive uses outside the standard lexicon and is particularly common in words pertaining to the register of abusive language. Here the voiceless alveolar affricate sometimes generates euphemistic variants of vulgar expressions or words inadmissible in polite usage: /haʔ alla ~ alleč/ ‘by Jove’, /il-madonna ~ il-madonneč/ ‘by our Lady’, etc. One of the most common terms of abuse in M is /ču:č/ (pl. /čwɪ:č/) ‘jackass!’ < It. *ciuco* ‘donkey’.

16.2.6. *Interference*

Throughout Malta's colonial period under British rule (1800–1964), knowledge of English was the key to social preferment and professional advancement. Many Maltese are bilingual in M and English; the average M speaker's command of English can range from fairly rudimentary to virtually native competence. M speakers also habitually codeswitch very freely between the two languages so that a M utterance is as likely as not to display embedded English lexical components—sometimes with minimal formal integration of the loaned elements to the native M sound pattern. Thus the surface content of a M utterance is quite liable to include nonnative sound units such as [æ], [ə], [θ], and [ð], though several fully integrated English words that have been adapted to M phonology are normally pronounced in their merged form even by educated M speakers: /kitla/ 'kettle', /tayer/ 'tire', /praymos/ 'primus stove', /trakk/ 'truck', etc. M speakers who habitually speak English at home sometimes carry over certain English phonetic features into their M speech, e.g., the alveolar articulation of M /t/ and /d/.

M speakers are also very familiar with Italian, but though M has integrated to different degrees a very large number of Italian words, intake of new Italianisms into M appears to be minimal, and habitual codeswitching between M and Italian nonexistent.

16.3. The Vowels

16.3.1. *Short Vowels*

SM has the system of five short vowels shown in Table 16-2.

Table 16-2. Maltese Short Vowels

i	u
e	o
a	

This contrasts with the situation obtaining in many rural M dialects which have a four-vowel system (Puech 1994: 26f.). The functional status of these short M vowels can be illustrated by means of the following word pairs:

<i>i : e</i>	<i>nisa</i>	'women'	<i>nesa</i>	'he forgot'
<i>i : a</i>	<i>wisa</i>	'breadth'	<i>wasal</i>	'he arrived'

<i>i : o</i>	<i>ġibna</i>	'we brought'	<i>ġobna</i>	'piece of cheese'
<i>i : u</i>	<i>siwi</i>	'value'	<i>suwed</i>	'black (pl.)'
<i>e : a</i>	<i>mera</i>	'mirror'	<i>mara</i>	'woman'
<i>e : o</i>	<i>ġenna</i>	'heaven'	<i>ġonna</i>	'gardens'
<i>e : u</i>	<i>dewwa</i>	'he healed'	<i>duwwa</i>	'medicine'
<i>a : o</i>	<i>sakra</i>	'drunkenness'	<i>sokra</i>	'lock'
<i>a : u</i>	<i>dawwar</i>	'he turned'	<i>duwwa</i>	'medicine'
<i>o : u</i>	<i>lottu</i>	'lotto'	<i>luttu</i>	'mourning'

As can be inferred from some of these pairs, vocalic contrasts involving /u/ are systemically weak in SM. Though /u/ is formally part of the M short vowel system—note its fairly widespread occurrence in unstressed syllables: /ġura:t/ 'grasshopper', /summɪ:n/ 'quails', /ġurɪ:n/ 'rat', etc.—it is virtually nonexistent in stressed syllables of lineally inherited lexicon, where it has undergone a systematic shift to /o/: /omm/ 'mother', /forn/ 'oven', etc. Many M words of Italian provenance have also undergone this shift: /ʃott/ 'dry' < *asciutto*, /ponn/ 'fist' < *pugno*, /dlonk/ 'frequently' < *di lungo*, etc. The low functional yield of the *u : o* contrast in SM misled Cohen (1970 [1966]: 140) into assigning the vowel [u] purely allophonic status in the SM sound system. However, there can be little doubt that the occurrence of stressed [u] in several well-integrated Italian terms of a learned nature and in certain recent loans from English justifies the assigning of full functional status to short stressed /u/:

<i>luttu</i>	'mourning'
<i>lussu</i>	'luxury'
<i>pulptu</i>	'pulpit'
<i>bust</i>	'bust'
<i>multa</i>	'fine (n.)'
<i>ġust</i>	'fair'
<i>futbol</i>	'football'
<i>šutt</i> (< Eng. <i>shoot</i>)	'shot (football)'

The distributional restriction of native **u* to unstressed positions in SM has yielded a morphophonemic vowel alternation of /o/ ⇒ /u/ concomitant with synchronic stress shift (e.g., in morphological derivation and proclisis), as in the following words:

<i>fost</i> (< <i>fi wast</i>)	'amid'	<i>fustá:ni</i>	'middle (adj.)'
<i>kóllu</i>	'all of it (masc.)'	<i>kull-hádd</i>	'everybody'

This alternation between stressed /o/ and pre-stress /u/ is often carried over into the morphophonology of well-integrated Romance lexicon:

<i>gost</i> (< Sic. <i>gustu</i>)	'delight'	<i>gustú:z</i>	'nice, handsome'
<i>yimpórta</i>	'it matters'	<i>ma yimpurtá:š</i>	'it doesn't matter'

There are other restrictions on short vowel distribution in SM. A fairly strict constraint in native words disallows the short high vowels /i/ and /u/ in prejunctural singly closed syllables, where only their mid counterparts, /e/ and /o/, respectively, can occur: *ra:ǵel* 'man', *yɪ:kol* 'he eats'. Many lexical loans also conform to this stricture:

<i>la:pes</i> (< It. <i>lapis</i>)	'pencil'
<i>so:der</i>	'soda (for laundry)'
<i>beyken</i>	'bacon'
<i>printes</i>	'apprentice'
<i>pro:set/pro:sit</i> (< It. <i>prosit</i> < Lat.)	'well done!'

Some, however, do not: /to:kis/ 'cinema'.

One consequence of the aforementioned general shift of stressed **u* to /o/ is that the phonological rule affecting underlying or historical high vowels in stem-final VC sequences yields an unsymmetrical surface distribution for underlying /i/ and /u/; thus, whereas underlying /i/ in stem-final VC sequences shows up as such in surface structure in non-final position, /u/ does not:

<i>nízzel</i>	'he brought down'
<i>nizzílom</i>	'he brought them down'
<i>šórob</i>	'he drank'
<i>šoróba</i> (not * <i>šorúba</i>)	'he drank it'

16.3.2. Long Vowels

M has retained the OA distinction between long and short vowels in open stressed syllables, though the functional load and present distribution of the length contrast in M are, to an extent, both products of its special internal history:

<i>nizel</i>	'he descended'	<i>ni:zel</i>	'descending (masc.)'
<i>ǵara</i>	'it happened'	<i>ǵa:ra</i>	'her neighbor'
<i>omma</i>	'her mother'	<i>o:mma</i> (< Ar. <i>gumma</i>)	'sadness'

Many speakers of SM automatically neutralize vowel length contrasts in unstressed positions:

<i>ferhá:n</i>	'happy (sg.)'	<i>ferhani:n</i>	'happy (pl.)'
<i>barrá:ni</i>	'foreign (masc.)'	<i>barraníyya</i>	'foreign (fem.)'

In view of the fact that short vowels in open syllables tend, as in many Arabic dialects, to be lost, most M final vowels in the native lexicon, except the

feminine ending {-a}, have an underlying length feature which is automatically lost in direct contact with word boundary, but is restored before consonantal suffixes:

<i>stí:dnú</i>	‘they invited’
<i>stidnú:h</i>	‘they invited him’
<i>ma stidnú:š</i>	‘they didn’t invite’

The function of the vocalic length contrast is minimal in lineally inherited monosyllables since words of this canonic shape tend to display a certain metrical equivalence between VCC and V:C (cf., for instance, the M stress rule in § 16.5 below):

<i>ġa:r</i>	‘neighbor’	<i>ġarr</i>	‘he carried’
<i>ra:s</i>	‘head’	<i>rass</i>	‘he tightened’

However, as already noted, the fusion of certain OA consonants with neighboring vowels in M has in some cases introduced new extra-long syllables of the type that are disallowed in many contemporary dialects: /e:dded/ ‘he threatened’ < OA *haddad*.

Vocalic length contrasts are also neutralized in monosyllables consisting of the canonic form CV:—a relatively rare type of monosyllable in M:

<i>ra(:)</i>	‘he saw’	
<i>ġi(:)</i>	‘he came’	
<i>zu(:)</i>	‘get out of the way!’	⇐ zu:l
<i>hu(:)</i>	‘take! (sg.)’	⇐ hu:d < Ar. <i>xud</i>

The etymologically long vowel in the underlying form |hu:d| represents the regular M reflex: /hu:du/ ‘take! (pl.)’ and /hu:dom/ ‘take them!’, OA **axað* and **akal* having been reconstructed as medially weak stems. The reflex of Ar. *la*: ‘no!’ is most often realized as *le(:)* or *le*”.

As in many Arabic dialects, primary vocalic length contrasts are also lost before a geminated consonant (mostly /ll/) in certain morphophonemic contexts:

<i>‘a:l</i>	‘he said’	
<i>‘alli</i>	‘he said to me’	
<i>zu:(l)</i>	‘Get out!’	
<i>zulli</i>	‘Get out of my way!’	
<i>kelli</i>	‘I had’	< * <i>ki:n</i> + <i>li:</i>
<i>ikollu</i>	‘he will have’	< * <i>iku:n</i> + <i>lu:</i>
<i>malli</i>	‘as soon as’	< * <i>ma’ illi</i>

Secondary vowel length in stressed syllables, however, tends to be retained before geminate consonants:

<i>ġa:nni</i>	‘he obliged me’	< * <i>ġa‘al</i> + <i>ni</i>
<i>a:mmed</i>	‘he baptized’	< * <i>‘ammad</i>

A striking feature of the SM long vowel system is its display of four degrees of aperture along the front of the vowel triangle as opposed to three along the back. It comprises the segments shown in Table 16-3.

Table 16-3. Maltese Long Vowels

i:	u:
ɪ:	
e:	o:
a:	

Functional contrasts between long vowels in SM can be illustrated by means of the following examples:

<i>i: : ɪ:</i>	<i>zi:d</i>	‘increase!’	<i>zɪ:d</i>	‘he increased’	
<i>i: : e:</i>	<i>dɪ:n</i>	‘this (f.)’	<i>de:n</i>	‘intelligence’	< <i>dihn</i>
<i>i: : a:</i>	<i>dɪ:n</i>	‘this (f.)’	<i>da:n</i>	‘this (masc.)’	
<i>i: : o:</i>	<i>sɪ:ba</i>	‘find her!’	<i>so:ba</i>	‘repentance’	~ <i>ʃ‘b</i>
<i>i: : u:</i>	<i>li:ra</i>	‘pound’	<i>lu:ra</i>	‘backwards’	< <i>l + wara:</i> ’
<i>ɪ: : e:</i>	<i>wɪ:d</i>	‘valley’	<i>we:da</i>	‘promise’	< * <i>w‘d</i>
<i>ɪ: : a:</i>	<i>sɪ:’</i>	‘leg’	<i>sa:’</i>	‘he drove’	
<i>ɪ: : o:</i>	<i>lɪ:ma</i>	‘which?’	<i>lo:ba</i>	‘game’	< <i>lu‘ba</i>
<i>ɪ: : u:</i>	<i>sɪ:’</i>	‘leg’	<i>su:’</i>	‘drive!’	
<i>e: : a:</i>	<i>de:r</i>	‘he appeared’	<i>da:r</i>	‘house’	
<i>e: : o:</i>	<i>re:ba</i>	‘greed’	<i>lo:ba</i>	‘game’	
<i>e: : u:</i>	<i>de:r</i>	‘he appeared’	<i>du:r</i>	‘turn round!’	
<i>a: : o:</i>	<i>na:l</i>	‘horseshoe’	<i>no:l</i>	‘freight’	< It. <i>nolo</i>
<i>a: : u:</i>	<i>da:m</i>	‘he lingered’	<i>du:m</i>	‘take your time!’	

Whereas the oppositions between the SM long vowels /i:/, /ɪ:/, and /u:/ continue historical contrasts between OA long vowels, the SM mid vowels /e:/ and /o:/ do not derive, as in most Arabic dialects, from /ay/ and /aw/, but usually represent the outcome of secondary developments, e.g., fusion of short **i* and **u* with adjacent velar, pharyngeal, and laryngeal fricatives:

SM	Meaning	Old Arabic
<i>ʃo:l</i>	‘work’	< <i>ʃuyl</i>
<i>so:la</i>	‘cough’	< <i>su‘la</i>
<i>de:n</i>	‘intelligence’	< * <i>dihn</i> < <i>ðihn</i>

Phonological analyses of M that postulate an “abstract *ghajn*” restrict the distribution of phonological length in vowels by implying that surface length in certain vowels can be interpreted as deriving from underlying bi-phonemic combinations of “*gh* + short vowel” (or vice versa).

The aforementioned asymmetrical patterning of long M vowels occasioned by the *ima:la*-shift presents a marked contrast with the situation obtaining among mainstream Arabic dialects showing a form of *ima:la šadi:da*, since these tend to fuse the resulting front vowel with the qualitatively closest available front vowel, i.e., with [e:] < **ay*, or with [i:].

The systemic asymmetry obtaining among the long vowels of SM naturally also extends to systemic relationships between long and short segments. Thus the morphophonemic lax correlate of destressed SM /ɪ:/ is traditionally /e/ rather than /i/:

<i>bní:dem</i>	‘man’	<i>bnedmí:n</i>	‘men’
<i>í:ni</i> (< <i>ha:ni</i>)	‘glad’	<i>enyí:n</i>	‘glad (pl.)’
<i>mí:ra</i>	‘he contradicted’	<i>merí:ni</i>	‘he contradicted me’
<i>wí:hed</i>	‘one’	<i>wehída</i>	‘alone (f.)’
<i>kí:fer</i>	‘cruel’	<i>kefrí:n</i>	‘cruel (pl.)’

Among SM speakers, however, this is a residual trait; thus outside specific morphophonemic or lexical contexts, the short vowel /i/ is by far the most common surface realization of a reduced /ɪ:/ in SM: /nɪ:zel/, pl. /nizli:n/ ‘descending’, /rɪ:’ed/, pl. /ri’di:n/ ‘sleeping’, etc. Lexicalization of the lax correlate of */ɪ:/ has occurred under stress in M /kell-/ ‘to have’ < *[kɪ:n + l-]: /kelli/ ‘I had’, where it is probably conditioned by the following cluster.

To nonnative ears the auditory difference between SM /i:/ and /ɪ:/ appears to be a very slight one. The phonetic quality of SM /i:/ closely approximates that of the vowel in Eng. *bead*, whereas the *ima:la* reflex is much closer to long Eng. [ɪ:] occurring before the voiced stops, e.g., *bid* [bɪ:d]. The functional contrast between SM /i:/ and /ɪ:/ carries a significant functional load supported by several minimal pairs of high frequency:

<i>smi:n</i>	‘fat (masc.sg.)’	<i>smɪ:n</i>	‘fat (pl.)’
<i>sni:n</i>	‘years’	<i>snɪ:n</i>	‘teeth’
<i>ǧi:b</i>	‘bring!’	<i>ǧɪ:b</i>	‘he brought’
<i>ʔi:s</i>	‘measure!’	<i>ʔɪ:s</i>	‘he measured’

In SM the historical distinction between OA /i:/ and the *ima:la* reflex is neutralized before /ʔ/ (< **q*) and /h/ (< **h*, **x*, and **h*) where only [ɪ:] occurs:

<i>sɪ:ʔ</i>	‘leg’	< <i>sa:q</i>
<i>drɪ:a</i>	‘narrowness’	< <i>ǧi:qa</i>

<i>ġi:h</i>	‘honor’	< <i>ġa:h</i>
<i>fi:h</i>	‘in him’	< <i>fi:h</i>

Many SM speakers (including the present writer) also neutralize this contrast in certain morphological contexts, e.g., in the first person plural inflectional suffix {-na:} before object suffixes {-k} and {-kom}:
ġibna ‘we brought’
ġibni:k (~ -ni:k) ‘we brought you (sg.)’
ġibni:kom (~ -ni:kom) ‘we brought you (pl.)’

16.3.3. Diphthongs

SM has the seven diphthongs shown in Table 16-4.

Table 16-4. Maltese Diphthongs

iw	ew	aw	ow
–	ey	ay	oy

The systemic asymmetry within the class of M diphthongs derives from the fact that SM /i:/ is usually realized as a monophthong rather than as a diphthongized [iy].

The diphthong /iw/ is highly marked in SM, no doubt on account of the already mentioned articulatory incompatibility of lip-spreading and lip-rounding entailed. This diphthong appears to be restricted to the term /liw-ya/ ‘bend’, and Aquilina (1987: 742) significantly notes the dialectal variant *lilwa* showing the shift of the problematic bilabial glide to the second syllable. Secondary /iw/ arising in morphological inflection tends to show up as surface [u:], e.g., in [u:sa] ‘wider’ ⇐ liwsa:l < Ar. *awsa*‘, the OA *af’al* pattern generally yielding M *if’a/el*: /ikbar ~ akbar/ ‘bigger (masc.)’, /isbah/ ‘more beautiful (masc.)’, /içken/ ‘smaller (masc.)’, /iswed/ ‘black (masc.)’, etc. The proliferation of new diphthongs in SM was principally a historical consequence of loss of emphasis:

<i>newl</i>	‘loom’	< <i>nawl</i>
<i>dawl</i>	‘light (n.)’	< <i>daw</i> ’
<i>seyf</i>	‘dagger’	< <i>sayf</i>
<i>sayf</i>	‘summer’	< <i>şayf</i>

In SM, the first element in the diphthong /ew/ can be optionally rounded and backed in contact with the bilabial glide /w/, yielding free variants such as:

<i>lewn</i> ~ <i>lown</i>	‘color’
<i>bewl</i> ~ <i>bowl</i>	‘urine’

Note, for instance, the common SM realization of the English loanword 'bowl' as /bawl/ avoiding homophony with the M variant [bowl] 'urine'.

SM primary and secondary diphthongs continuing OA /aw/, /ay/, /i:/, and /u:/ after historical /-/ tend to be synchronically marked in M (especially among conservative speakers) and are usually phonetically longer than other diphthongs: [a:yn] 'eye' < Ar. *ʿayn*, but /tayn/ 'mud' < Ar. *ṭīn*. SM also has the diphthong /oy/, e.g., /voyt/ 'empty' < Sic. *voitu*, /boyy/ 'hey you!' < Eng. *boy*.

16.3.4. *The Synchronic ima:la*

The aforementioned *ima:la*-shift in M, a different variety of which was recorded for Mesopotamian Arabic in *Kita:b* by Sibawayhi (died c. 800), presents the Arabist with what is possibly the most striking areal trait in this otherwise highly deviant offshoot of vernacular Arabic. The phonetic realizations of the *ima:la*-vowel in M vary considerably from one dialect to the next (cf. the folktexts in nonstandard M reproduced in Stumme 1904, and in Puech 1994). In SM *ima:la*-shift affects stressed long *[a:], usually, though not exclusively, in non-emphatic or non-back consonantal environments yielding, for instance, /ktɪ:b/ 'book', /tiġɪ:ġa/ 'hen' < **kta:b* and **da/iġa:ġa*, as opposed to /sa:r/ 'he became', /sa:/ 'he drove' < *ša:r* and *sa:q*. In unstressed position, the M reflex of *[a:] retains the [low] feature: /ílma/ 'water' < **il-má:*; /hobla/ 'pregnant' < *ḥubla:*; /mela/ 'therefore' (~ /immela/) < **imma:la*: (Dozy 1881, 1:36), /ébdal/ 'none' < **ʿabda*: < *ʿabadan* 'never', etc.

The mutual historical ordering obtaining between the stress-assignment rule and the *ima:la*-shift observable in Medieval Maltese has been integrated into the synchronic grammar of the language, where it has contributed a certain degree of formal complexity to the morphophonemics of verb inflection and suffixation.

In SM the synchronic *ima:la*-rule affects underlying long /a:/ chiefly: (i) in the stems of many finally weak verbs, (ii) in the subject and object pronominal suffix {-na:}, and (iii) in the 3rd person feminine suffix {-ha:}. Observe, for instance, the derivations of surface forms from underlying /fela: + hl 'he scrutinized him' and /ráyna: + hl 'we saw him':

(1) The synchronic *ima:la*-rule in SM

(a) Underlying form	<i>féla:</i> + <i>h</i>	'he scrutinized him'
Stress shift	<i>felá:h</i>	
<i>Ima:la</i> -rule	<i>felí:h</i>	
Syncope	<i>flí:h</i>	

(b) Underlying form	<i>ráyna: + h</i>	‘we saw him’
Stress shift	<i>rayná:h</i>	
<i>Ima:la</i> -rule	<i>rayní:h</i>	

The circumstance that only synchronically stressed occurrences of SM /a:/ are affected by the *ima:la*-rule necessarily implies that grammatically complex strings displaying more than one instance of the low central vowel /a:/ (e.g., concomitant with two or three layers of person markers) cannot be simply derived via fronting the last /a:/ in the string, but require recurring application of the rule of stress assignment and of the *ima:la*-rule, in that order. A representative SM example of such a string is the verb phrase like <ma sraqnihliex> ‘we did not steal it from her’, which would seem to derive morphophonemically via cyclical application of the stress and *ima:la* rules to an underlying representation such as the following string: [ma [[[srá:na:]ha:]lha:]]š] (cf. Brame 1974).

16.3.4.1. Morphophonemic *ima:la* in Romance Quadriliteral Plurals

The M synchronic *ima:la* shows up in another morphophonemic process: broken plural formation on the M scheme CCV:CVC (< Maghrebi Ar. *f'a:lel*) for many quadriliteral roots which Maltese has derived from Italian singular forms.

Plural forms in this noun class select /a:/ or /ɪ:/ in accordance with a rule matching the advancement feature of the vowel in the plural with that of the stressed vowel in the singular. In other words, a stressed /i/ or /e/ in the singular form requires the *ima:la* vowel /ɪ:/ in the plural, whereas stressed non-front vowels (e.g., /á:/ or /ú:/) require /a:/:

Singular		Plural
<i>ber[r]ítta</i>	‘cap’	<i>brí:ret</i>
<i>furkétta</i>	‘fork’	<i>frí:ket</i>
<i>čavétta</i>	‘key’	<i>čwí:vet</i>
<i>kappéll</i>	‘hat’	<i>kpí:pel</i>
<i>bastú:n</i>	‘walking stick’	<i>bsa:ten</i>
<i>kaššú:n</i>	‘drawer’	<i>kša:šen</i>
<i>pastá:s</i>	‘rude’	<i>psa:tas</i>
<i>fardá:l</i>	‘apron’	<i>fra:dal</i>

As can be inferred from the last four examples, singular stems displaying the low stressed /á:/ require the low vowel /a/ in the final syllable of the plural, whereas singular stems displaying a non-low vowel (e.g., /i/ or /u:/) take /e/.

The morphophonemic rule assigning the *ima:la* vowel in quadriliteral plurals of Romance singular nouns was presumably extrapolated by M

speakers on the analogical basis of a perceived relationship between front and non-front vowels in certain frequently used quadriradicals pertaining to the inherited lexicon:

Singular		Plural
<i>zardu:na</i>	'shoe'	<i>zra:ben</i>
<i>sikki:na</i>	'knife'	<i>skɪ:ken</i>
<i>farta:s</i>	'a bald man'	<i>fra:tas</i>

16.4. Phonotactics

16.4.1. Consonant Clusters

Whereas consonant clusters in the native lexicon generally tend to enjoy stability both idiolectally and across the M speech community, those pertaining to the Romance lexicon are often susceptible to variation conditioned by stylistic and other factors, e.g., Low /stu:ptu/ 'stupid', High /stu:pidu/ (< It. *stupido*).

The nature and distribution of initial clusters in the inherited lexicon of M are closely analogous to those obtaining in many Arabic dialects (Cantineau's *parlers non différentiels*) that generally elide all short vowels in unstressed open syllables: /klɪ:b/ 'dogs', /trɪ:/ 'road', /tra:b/ 'dust', /nifthu/ 'we open', /yiksru/ 'they break', etc. Many well-integrated Italian loans in M have also been adapted to these syllabic norms: /zgu:r/ 'sure' < *sicuro*, /lvant/ 'east' < *levante*, /lment/ 'complaint' < *lamento*, /kampna:r/ 'bell tower' < Sic. *campanaru*, but the effect of this acculturation process has not introduced radically novel clustering patterns in M.

In both native and borrowed lexicon, initial CC clusters beginning with a resonant require a prosthetic vowel after pause, [ilsɪ:n] 'tongue', [imwéyyed] 'tables', [ilvánt] 'east', [irvéll] 'disturbance', etc. The same goes for initial geminate clusters, which are particularly common in M denominal verbs of Italian origin: /irra:bya/ 'he became angry' ⇐ /ra:bya/ 'anger', /ikko:pya/ 'he copied' ⇐ /ko:pya/ 'copy', /ippri:tka/ 'he preached' ⇐ /pri:tka/ 'sermon', etc.

The occurrence of the definite article {(i)l-} before initial CC clusters does not necessarily trigger off an auxiliary vowel, but Italian clusters with initial sibilants do: /il-kbi:r/ 'the big one', /il-hrɪ:f/ 'the lambs', but /l-isko:la/, /l-ispera:g/ 'the string', /l-izball/ 'the mistake'. In relaxed speech, however, initial CC clusters with a resonant in first position often dispense with the secondary syllabic and show assimilation of the article: /ir-rmɪ:d/ 'the ashes'. In certain words, the secondary prosthetic vowel has become lexicalized and

appears exclusively after the article even though the initial CC cluster has been dissolved, e.g., via loss of historical *ghajn*:

<i>ma:š</i>	‘interest’	< * <i>m’a:š</i>	< <i>ma’a:š</i>
<i>ma:rfa</i>	‘spoon’	< * <i>mɣarfa</i>	< * <i>mayárfa</i> < OA <i>mayrafa</i>
<i>l-ima:š</i>	‘the interest’	< * <i>l-im’a:š</i>	
<i>l-ima:rfa</i>	‘the spoon’	< * <i>l-imyarfa</i>	

Free variation between /l/ and /r/ before a consonant in new syllables, as in /ilsir ~ risir/ ‘slave’ < OA *yasi:r* should probably be accounted for within the explanatory framework outlined in § 16.2.4.7 for optimal positions of liquids in the syllable. Preference for /r/ as syllable coda is also evident from historical shifts like M /irkant/ ‘auction’ < It. *incanto*.

The synchronic treatment of medial CCC clusters in the inherited lexicon conforms to patterns familiar from other Arabic vernaculars, which generally show CVCC if medial C is a resonant: /yahílbu/ ‘they milk’, /yo’órbu/ ‘they approach’, /yo’ómsu/ ‘they hop’, etc. The accent regularly falls on the secondary syllabic.

Where no medial resonants are involved in potential CCC clusters, M is strikingly tolerant to medial obstruent clusters and displays some highly marked phonetic sequences especially in onomatopoeic local creations: /ipačpču/ ‘they chatter’, /ipa’p’u/ ‘they honk (car horn)’, /ihašwšu/ ‘they rustle (silk, paper)’, /i’as’su/ ‘they cut out’.

Underlying word-final CC clusters with a final resonant are resolved in surface structure through anaptyctic vowels, as in many Arabic dialects: /iben/ ‘son’, but /ibni/ ‘my son’. Word-final clusters in surface and underlying structure are rare in Italian loans, but note the synchronic elision of the final apical trill in M /surmast/ ⇐ |surmastr| ‘schoolmaster’, pl. /surmastr-iyɪ:t/.

16.4.2. Vowels in Hiatus

SM word phonology disallows the occurrence of vowels in immediate adjacency, i.e., without an intervening consonant. Thus, for instance, historical loss of intervocalic *//, */ɣ/, or */h/ invariably yields a secondary bilabial or palatal glide agreeing in advancement and lip-rounding with the preceding stressed vowel:

(2) Diachronic treatment of hiatus in SM

OA			Orthography	
<i>šana:‘a</i>	> * <i>šni:‘a</i>	> <i>šniya</i>	<i>xniegħa</i>	‘rumor’
<i>furu:ya</i>	> * <i>fru:ya</i>	> <i>fruwa</i>	<i>frughā</i>	‘emptiness’
* <i>hubu:la</i>	> * <i>blu:ha</i>	> <i>bluwa</i>	<i>bluha</i>	‘foolishness’

Closely analogous morphophonemic treatment of elidable /h/ intervocalically can be observed in the suffixation of the object pronouns {-h}, {-ha:}, and {-hom} to the stems of finally weak verbs. Note, for instance, the synchronic derivations of SM <ksieha> 'he covered it (fem.)', and SM <jixtru-hom> 'they buy them', here presented in condensed fashion:

- (3) Elimination of hiatus in verb stems with object pronouns
- | | | | |
|-----|-----------------------|----------------------|------------------------|
| (a) | Underlying form | <i>késa: + ha:</i> | 'he covered it (fem.)' |
| | Stress shift | <i>kesá:ha</i> | |
| | <i>Ima:la</i> -rule | <i>kesí:ha</i> | |
| | Syncope | <i>ksí:ha</i> | |
| | Deletion of laryngeal | <i>ksíya</i> | |
| (b) | Underlying form | <i>yíštru: + hom</i> | 'they buy them' |
| | Stress shift | <i>yíštrú:hom</i> | |
| | Deletion of laryngeal | <i>yíštrúwom</i> | |

As can be inferred from these two derivations, stressed stem-final high vowels undergo the following changes before the suffixed pronouns /-ha/ and /-hom/:

<i>l-í:h-l</i>	$\Rightarrow$ <i>l-i:y-l</i>	$\Rightarrow$ /-iy-/
<i>l-ú:h-l</i>	$\Rightarrow$ <i>l-u:w-l</i>	$\Rightarrow$ /-uw-/

which recapitulate the historical sound shift we observed word-internally in (2). Vocalic length reduction here can be plausibly attributed to the phonologically weak nature of the syllabic arrest provided by semivowels after long high vowels. It is not without interest in this connection that certain non-standard M dialects (e.g., that spoken in Nadur, Gozo) display secondary gemination of the post-stress bilabial glide with concomitant reduction of word-final *l-u:l* in verbs with pronominal suffixes: /triduwwim/ 'you want them' (Puech 1994: 87) < **tri:du:him*. Observe also the diachronic tendency in SM towards gemination of historical /y/ after a stressed long high vowel in reflexes of OA *fa'a:lil* and *fa'a:li:l* plurals:

M

<i>kneyyes</i>	< * <i>kne:yis</i>	< <i>kana:yis</i>	'churches'
<i>mweyyed</i>	< * <i>mwe:yid</i>	< <i>mawa:yid</i>	'tables'

On the other hand, stems which retain the low reflex of final /a:/ in the context of syntactic suffixes (e.g., /ra:/ 'he saw', /štara/ 'he bought', < **ra:* and **štara:*) circumvent vocalic hiatus differently. M morphophonology here resorts to syllabic fusion, which is, as far as one can tell from the Arabic dialectal literature, unique to M. In its simplest form one can see it operating in the sequences <raha> 'he saw her' and <xtraha> 'he bought it (fem.)', both of

which delete the intervocalic laryngeal and yield the monosyllables [ra(:)] and [štra(:)] respectively, with optional vowel length. A more extreme form of syllable fusion occurs in verb phrases with the object pronoun {-hom}, as in realizations of M <rahom> ‘he saw them’ [ro:m] and <xtrahom> ‘he bought them’ [štro:m]:

- (4) Elimination of vocalic hiatus via syllable fusion
- | | | | |
|-----|-----------------------|-------------------|------------------|
| (a) | Underlying form | <i>rá:+ hom</i> | ‘he saw them’ |
| | Deletion of laryngeal | <i>rá:om</i> | |
| | Syllabic fusion | <i>ró:m</i> | |
| (b) | Underlying form | <i>štára:+hom</i> | ‘he bought them’ |
| | Stress shift | <i>štrá:hom</i> | |
| | Deletion of laryngeal | <i>štrá:om</i> | |
| | Syllabic fusion | <i>štró:m</i> | |

These schematic presentations illustrate how M systematically eliminates hiatus present in the underlying form of verb phrases and generates surface structures that conform to the phonetic strictures of M word phonology. There are, however, cases where unique solutions are not available for smoothing out potentially abrupt transitions in underlying structures. One such case occurs in finally weak verbs with a morphophonemically stable stem-final /a:/, i.e., one that is not affected by the synchronic *ima:la*-rule. When such stems display a double object: <xtrahomli> ‘he bought them for me’, <xtrahieli> ‘he bought it for me’, etc., word stress shifts to the direct object suffix yielding forms like [štra.ómlɪ] and [štra.í:li]), etc., where the height and advancement discrepancies between the vowels flanking the syllable boundary render the hiatus highly marked. Many M speakers often simply retain the problematic hiatus. In relaxed speech, however, they often resort to different ways of smoothing the abrupt transition, e.g., by introducing the glide [y]: [štrayí:li]; or a geminated [h] on the analogy of stems displaying final abstract *ghajn*: [štraħħómlɪ] (cf. [refaħħómlɪ] ‘he carried them for me’).

16.4.3. Roundness Assimilation

Unrounded short vowels in the inherited lexicon occur very freely in adjacent syllables:

<i>tifel</i>	‘boy’	<i>siġar</i>	‘trees’
<i>wera’</i>	‘leaves’	<i>hanek</i>	‘gums’

By contrast, the distribution of the rounded mid back vowel /o/ (but not /u/) is subject to a rule of progressive roundness assimilation both intra- and intermorphemically:

<i>to'ol</i>	'weight'	<i>hodon</i>	'bosom'
<i>borom</i>	'cooking pots'	<i>torok</i>	'Turks'
<i>šorob</i>	'he drank'	<i>romol</i>	'he became a widower'

Across word boundary, the assimilation rule affects the allomorphy of the 2nd person suffix *-Vkl*. There is a certain amount of dialectal variation in the phonemic realization of this suffix. In SM three allomorphs occur, i.e., {-ok} after stems with /o/, {-ak} after historically stem-final //, and {-ek} elsewhere:

<i>omm-ok</i>	'your mother'	<i>tiy-ak</i>	'your'
<i>oht-ok</i>	'your sister'	<i>ra:s-ek</i>	'your head'
<i>hoġr-ok</i>	'your lap'	<i>si:ʔ-ek</i>	'your leg'
<i>flo:k-ok</i>	'instead of you' (< Ar. * <i>fī</i> + Sic. <i>locu</i>)		
<i>post-ok</i>	'your place' (cf. Sic. <i>postu</i>)		

When {-Vk} is suffixed to verb stems, the "harmony rule" operates on underlying representations and copies the roundness feature of stem vowels that are subsequently elided at surface level:

Underlying form	<i>yɪ:hod + Vkl</i>	'he takes you'
Roundness assimilation	<i>yɪ:hodok</i>	
Syncope	<i>yɪ:hdok</i>	

Further examples are: /*yigbrok*/ 'he picks you up' ⇐ /*yigbor + Vkl*/, /*yɪ:klok*/ 'he eats you up' ⇐ /*yɪ:kol + Vkl*/.

A less systematic, mostly optional, form of roundness assimilation (usually affecting /i/) also occurs in the Romance lexicon of M in the neighborhood of /u/ where it affects short unstressed vowels in the adjacency of /u/ bearing primary and secondary stress on: M /*pùlicíyya* ~ *pùlucíyya*/ 'police' < Sic. *pulizzia*, /*kùritú:r* ~ *kùrutú:r*/ 'corridor' < Sic. *currituri*, /*pruzuntu:z*/ 'rash' < Sic. *presuntusu*.

16.5. Stress

The assignment of word stress in native M words and in fully integrated loans is generally predictable by reference to their syllabic or morphemic composition. Like many Arabic dialects, M generally accentuates words without syntactical affixes on the VCC or V:C sequence closest to word boundary; in the absence of a VCC and V:C sequence, stress falls on penultimate syllables:

M	<i>háregħ</i>	'he went out'
	<i>wí:hed</i>	'one (masc.)'

<i>hallí:l</i>	‘thief’
<i>tuffí:ha</i>	‘apple’
<i>gabillótt</i>	‘farmer’
<i>purtiná:r</i>	‘janitor’

M shows very extensive historical loss of short vowels in unstressed open syllables; this factor has tended to restrict canonic forms in the native lexicon and in well-integrated foreign words to mono- or bi-syllabic shapes:

<i>kti:b</i>	‘book’	< <i>kita:b</i>
<i>slɪ:m</i>	‘peace’	< <i>sala:m</i>
<i>tra:b</i>	‘dust’	< <i>tura:b</i>
<i>mizbla</i>	‘dump’	< <i>mazbala</i>
<i>mahfra</i>	‘forgiveness’	< <i>mayfira</i>

Consequently most M words show stress on final or penultimate syllables. In synchronic terms, this means that when morphological inflection introduces additional syllables into underlying structure, morphophonemic adjustments occur in the surface form promoting closed syllabicity in unaccented positions and reducing the stem to one of the standard syllabic templates: *raʿad* + *u* ⇒ /raʿdu/ ‘they slept’; *dendel* + *u* ⇒ /dendlu/ ‘they hung’, *habbar* + *et* ⇒ /habbret/ ‘she announced’, etc. M words of Italian provenance can vary considerably in length and syllabic structure, depending on the degree to which they have been formally adapted to native canonic shapes. Massive influx of Romance stems into M over a long period of time has introduced atypical canonic forms into the language, entailing greater variability in word length, an enhanced tolerance to open syllabicity, a freer and richer distribution of vowels often contravening inherited vocalic schemes (particularly in unstressed syllables), etc.

The differential treatment meted out to Italian lexical material in M (no doubt correlating with the diffusional path selected by the lexical item in question) can be exemplified in the following words which range from virtually “whole” Italian, like (a), to more fully “merged” Italian, like (c), which has been radically restructured to conform to the native class of Form II verbs (/kisser/ ‘he broke’, /nizzel/ ‘he brought down’, etc.):

- | | | |
|------------------------|--------------|--------------------------------|
| (a) <i>edukacyo:ni</i> | ‘education’ | < It. <i>educazione</i> |
| (b) <i>peri:t</i> | ‘architect’ | < Sic. <i>peritu</i> |
| (c) <i>pitter</i> | ‘he painted’ | ⇐ {p-t-r} < It. <i>pittore</i> |

Educated M speakers freely resort to Italian words of a learned nature which tend to contravene the canonic forms of M and show atypical stress patterns, e.g., Italian proparoxytones (*sdrucchiole*), and words with final stress (*parole tronche*):

<i>píllola</i> ~ <i>pínnola</i>	'pill'	< It. <i>pillola</i> , Sic. <i>pinnula</i>
<i>dé:boli</i> ~ <i>débboli</i>	'weak'	< It. <i>debole</i> , Sic. <i>debbuli</i>
<i>virtú</i>	'virtue'	< It. <i>virtù</i>

Interestingly, in reflexes of certain recently borrowed Italian proparoxytones M speakers tend to hesitate between a native stress pattern and the Italian one: /ferti:li ~ fértili/ 'fertile' < It. [fértili]. However, most speakers of M have fully integrated the proparoxytonic accentual pattern via English loans like /šárabank/ 'bus', /télefown/, 'telephone'. As can be inferred from M /ferti:li/ here, M has phonologized the automatic phonetic length of Italian vowels in open syllables, and this often serves as the basis of stress assignment. That the integration in M of Italian words with open syllables at two removes from stress was a gradual process can be inferred from a lexical stratum showing secondary consonant gemination in this position:

<i>annima:l</i>	'animal'	< It. <i>animale</i>
-----------------	----------	----------------------

The integration of these phonologically long canonic forms in M with concomitant retention of pretonic short vowels, can be said to have stabilized in M the category of phonetic "secondary stress" in isolated words.

Though stress assignment is highly predictable in M, even where morphological factors are involved, implementation of the stress rule can still be relatively complex since it invariably interacts with a number of phonological rules and constraints: short vowel elision, raising and fronting of /a:/, reduction of underlying length, etc. Brame (1974: 56), for instance, noted that the M stress assignment rule systematically discriminates between morphological and syntactic affixes on verb stems, as can be seen from the following surface forms of the M verb /hareğ/:

lháreğ + nal	'we went out'	⇒ <i>hriğna</i>
lháreğ + nal	'he took us out'	⇒ <i>hariğna</i>

where the open syllable is elided in the first example but not in the second. Interaction between the stress assignment rule and morphophonemic processes can be particularly complex in the case of phonological words deriving from syntactic strings, as already noted in the section on *ima:la* (§ 16.4.4).

16.6. Reduction of the *nisba*-ending {-íyy}

As in many mainstream Arabic vernaculars, the singular forms of the OA *nisba* {-íyy} yield the following reflexes in M: masc. -i, fem. -iyya, where the masculine form has undergone length neutralization concomitant with the diachronic shift of stress to the stem vowel:

masc.	fem.	
<i>ma:l̥ti</i>	<i>maltiyya</i>	‘Maltese’
<i>a:rbi</i>	<i>arbiyya</i>	‘Arab’
<i>ramli</i>	<i>ramliyya</i>	‘sandy’

In the plural, the M form of the *nisba* for both genders has the invariant shape {-i:n} through haplology of the bisyllabic underlying and historical form **iyyi:n*:

<i>malti:n</i>	‘Maltese people’
<i>s’alli:n</i>	‘Sicilians’
<i>belti:n</i>	‘people from Valletta’

16.7. A note on phoneme–grapheme correspondences in Maltese

The present M orthography, standardized in 1934, is the product of a long but intermittent literary tradition harking back at least to the end of the 15th century, when the earliest extant text in M was composed: *Peter Caxaro’s Cantilena* (see Wettinger and Fsadni 1968). (The only other medieval texts discovered so far in the Maltese Islands are in a local form of Judeo-Arabic written in Hebrew characters; cf. Wettinger 1985).

However, it is only from the 18th century onwards that M has been written with any degree of continuity, though descriptive accounts of the language, necessarily entailing attempts at orthographic representation, were compiled before then: for instance, the recently discovered *Regole per la lingua maltese*, which is thought to date back to the late 17th century (cf. Casola 1988: 63).

In view of the present lack of a modern systematic description of M in languages other than M itself, it may be useful to conclude this sketch of M phonology with some general indications as to how the letters of the M alphabet correlate with the sound system of the language, if only to enable the interested linguist to avail himself of printed texts in M.

The M orthography presents a number of pitfalls for those unfamiliar with the spoken language. Thus except for the digraph *ie*, which is generally realized long, M does not ordinarily mark vocalic length. The circumflex accent is sometimes used to avoid ambiguity: *qartas* ‘to wrap in paper’ as distinct from *qartâs* ‘a paper funnel for wrapping groceries’ (Aquilina 1990: 1134).

The distribution of diacritics differentiating between sounds represented by the same letter does not always follow generally accepted linguistic prin-

ciples, e.g., restriction of a diacritic to the phonologically marked or statistically less frequent segment. Thus M oddly marks its native voiced alveolar fricative /z/ (< OA z) with a superscript dot <ž> to distinguish it from the alveolar affricate /c/ (< It. z) though the opposite procedure would have been more appropriate.

16.8. The Maltese alphabet

The letters of the M alphabet are here listed in the order usually adhered to in dictionaries; these are followed by their traditional names in square brackets, a phonetic description, and a few examples. (The digraph <ie> is not included as a separate symbol in the M alphabet.)

- a, A* [a:]—a low central vowel, long or short: *qam* ‘he rose’ [ˈa:m], *wasal* ‘he arrived’ [wásal]
b, B [be:]—a voiced bilabial stop: *bir* ‘well’ [bi:r]; *aħbar* ‘piece of news’ [aħbár]; *bieb* ‘door’ [bɪ:p]
ċ, Ċ [če:]—a voiceless alveolar affricate: *ċar* ‘clear’ [tʃa:r]; *keċċa* ‘he expelled’ [kétʃa]
d, D [de:]—a voiced dental stop: *dar* ‘house’ [da:r]; *beda* ‘he began’ [béda]
e, E [e:]—a mid front vowel, long or short: *erbgha* ‘four’ [é:rba]; *deheb* ‘gold’ [de:p]; *dell* ‘shadow’ [del]
f, F [éffe]—a voiceless labiodental fricative: *fiehem* ‘he explained’ [fíyem]; *siefer* ‘he traveled’ [sí:fer]; *ħlief* ‘except’ [ħlɪ:f]
ġ, Ġ [ġe:]—a voiced alveopalatal affricate: *ġar* ‘neighbor’ [ġa:r]; *riġel* ‘foot’ [ríġel]
g, G [ga:]—a voiced velar stop: *gżira* ‘island’ [gzi:ra]; *niggeż* ‘he pricked’ [nígges]
gh, Gh [a:yn]—a digraph usually continuing OA [ɨ] and [ɣ] yielding mostly vocalic length in stressed positions: *ghamel* ‘he made’ [a:mel], and [ħ] when closing word-final stressed syllables: *qiegh* ‘bottom’ [ɪ:ħ]
h, H [ákka]—a letter that continues etymologically the OA laryngeal fricative /h/; it is mostly realized as vocalic length: *hena* ‘joy’ [e:na]
ħ, H [éħhe]—a voiceless fricative with a widely variable point of articulation including velar, laryngeal, but mostly pharyngeal realizations: *ħafer* [háfer] ‘he forgave’; *baħar* ‘sea’ [báħar]; *riħ* ‘wind’ [rɪ:ħ]
i, I [i:]—a high front vowel, long or short: *iben* ‘son’ [íben]; *irid* ‘he wants’ [iri:t]

- j, J* [ye:]—a palatal glide: *jiena* 'I' [yí:na]; *sejjer* 'going (masc.sg.)' [séyyer]
- k, K* [ke:]—a voiceless velar stop: *kelb* 'dog' [kelp]; *beka* 'he wept' [béka]; *fenek* 'rabbit' [fének]
- l, L* [élle]—an alveolar lateral resonant: *lagħab* 'he played' [la:p]; *mielaħ* 'salty' [mí:laħ]; *dell* 'shadow' [del]
- m, M* [émme]—a bilabial nasal resonant: *mera* 'mirror' [méra]; *komma* 'sleeve' [kómma]; *tmiem* 'end' [tmí:m]
- n, N* [énne]—an alveolar nasal resonant: *nehħa* 'he took away' [néhħa]; *bnin* 'beneficial' [bni:n]
- o, O* [o:]—a mid back rounded vowel, long or short: *omm* 'mother' [ómm]; *għoli* 'high' [ó:li]
- p, P* [pe:]—a voiceless bilabial stop: *pastas* 'rude' [pastá:s]; *kappar* 'capers' [kappá:r]
- q, Q* [ʼa:]—a voiceless glottal stop: *qal* 'he said' [ʼa:l]; *qorti* 'law court' [ʼórti]; *triq* 'road' [trí:]
- r, R* [érre]—an apical trill: *raba* 'fields' [rába]; *mor* 'bitter' [mor]
- s, S* [ésse]—a voiceless alveolar fricative: *seraq* 'he stole' [séra]; *kesa* 'he covered' [késa]; *biss* 'only' [bis]
- t, T* [te:]—a voiceless dental stop: *tar* 'he flew' [ta:r]; *kattar* 'he increased' [káttar]
- u, U* [u:]—a rounded high back vowel, long or short: *usa* 'wider' [ú:sa]; *iżur* 'he visits' [izú:r]; *gurat* 'grasshopper' [ğurá:t]
- v, V* [ve:]—a voiced labiodental fricative: *venven* 'blow loudly (wind)' [vénnen]; *iva* 'yes' [i:va]
- w, W* [wa:w] or [we:]—*wera* 'he showed' [wéra]; *sewwa* 'he repaired' [sew-wa]; *raw* 'they saw' [ra:w]
- x, X* [éšše]—a voiced or voiceless alveopalatal fricative: *xorob* 'he drank' [šórop]; *televixin* 'television' [tèlevížin]
- ž, Ž* [za:]—a voiced alveolar fricative: *žar* 'he visited' [za:r]; *beža* 'he feared' [béza]
- z, Z* [ca:]—a voiced or voiceless alveolar affricate: *gazzetta* 'newspaper' [gaddzétta]; *zekzek* 'he tsked' [cékcek]; *gezzez* 'he bundled together' [géttssets]

Bibliography

- Al-Maghribi, 'Abd al-Qadir. 1947 [1908]. *Kitāb al iṣṭiqāq wa al-ta'rīb*, 2nd ed. Cairo: n.p.
- Aquilina, Joseph. 1987–90. *Maltese-English Dictionary*, 2 vols. Malta: Mid-sea Books.
- Barbato, A. 1990. "Conquista e presenza arabo-islamica a Malta." *Annali dell'Istituto Universitario Orientale*, Napoli 50: 233–47.
- Bauer, Leonhard. 1926 [1909]. *Das palästinische Arabisch*. Leipzig: Hinrichs.
- Blanc, Haim. 1953. *Studies in North Palestinian Arabic*. Jerusalem: Israel Oriental Society.
- . 1964. *Communal dialects in Baghdad*. Cambridge, Mass.: Harvard University Press.
- Borg, Alexander. 1976. "Reflexes of Pausal Forms in Maltese Rural Dialects?" *Israel Oriental Studies* 7: 211–25.
- . 1978. "A Historical and Comparative Phonology and Morphology of Maltese." Ph.D. dissertation, The Hebrew University, Jerusalem.
- . 1985. *Cypriot Arabic*. Stuttgart: Steiner.
- . 1994. "Observations on Some Evolutionary Parallels and Divergences in Cypriot Arabic and Maltese." *Actas del congreso internacional sobre interferencias lingüísticas árabo-romances y paralelos extra-iberos celebradas en Madrid del 10 al 14 de diciembre de 1990*, ed. Jordi Aguadé, Federico Corriente, and Marina Marugán, pp. 21–40. Zaragoza: n.p.
- Brame, Michael. 1972. "On the Abstractness of Phonology: Maltese ʕ." In *Contributions to Generative Phonology*, ed. M. K. Brame, pp. 22–61. Austin: University of Texas Press.
- . 1974. "The Cycle in Phonology: Stress in Palestinian, Maltese, and Spanish." *Linguistic Inquiry* 5: 39–60.
- Cassola, A. 1988. *Regole per la lingua Maltese*. Malta: Said International.
- Chejne, A. G. 1969. *The Arabic Language*. Minneapolis: University of Minnesota Press.
- Cohen, David. 1970 [1966]. "Le système phonologique du maltais." In *Études de linguistique sémitique et arabe*, pp. 126–49. The Hague: Mouton.
- Comrie, Bernard. 1980. "The Sun Letters in Maltese: Between Morphophonemics and Phonetics." *Studies in the Linguistic Sciences* 10 (2): 25–37.

- . 1986. "The Maltese Pharyngeal." *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39: 12–18.
- Dench, Geoffrey. 1975. *The Maltese in London*. London: Routledge and Kegan Paul.
- Dozy, Reinhart. 1881. *Supplément aux dictionnaires arabes*. Leiden: Brill.
- El-Hassan, S. A. 1977. "Educated Spoken Arabic in Egypt and the Levant: A Critical Review of Diglossia and Related Concepts." *Archivum Linguisticum* 9: 32–57.
- Ferguson, Charles A. 1959. "Diglossia." *Word* 15: 325–40.
- Firth, J. R. 1964. *The Tongues of Men and Speech* (Language and Language Learning 2). London: Oxford University Press. (Orig. publication 1937, 1930).
- Fischer, Wolfdietrich. 1959. *Die demonstrativen Bildungen der neuarabischen Dialekte*. The Hague: Mouton.
- . 1961. "Die Sprache der arabischen Sprachinsel in Uzbekistan." *Der Islam* 36: 232–63.
- , and Otto Jastrow, eds. 1980. *Handbuch der arabischen Dialekte*. Wiesbaden: Harrassowitz.
- Fleisch, Henri. 1974. *Études d'arabe dialectal*. Beirut: Dar al-Mashreq.
- Hull, Geoffrey. 1988. "Vicende e caratteristiche del maltese parlato in Egitto." *Incontri Siculo-Maltesi: Atti del convegno su Malta-Sicilia, Contiguità linguistica e culturale (Malta, 4–6 aprile 1986)*. Malta: University of Malta.
- Jeger, L. 1963. "London Maltese." *The Guardian*, Feb. 14.
- Kahane, Henry, and A. Tietze. 1958. *Lingua Franca in the Levant*. Urbana: University of Illinois Press.
- Mamo, J. 1930–31. *Ulled in-nanna Venut fl-Amerca* [Grandma Venut's children in America]. Malta: Author.
- Nachtigal, Gustav. 1974 [1879]. *Sahara and Sudan*. London: Hurst.
- Price, C. A. 1954. *Malta and the Maltese: A Study in Nineteenth Century Migration*. Melbourne: Georgia House.
- Puech, Gilbert. 1978. "A Cross-Dialectal Study of Vowel Harmony in Maltese." *Chicago Linguistic Society* 14: 377–89.
- . 1994. *Ethnotextes Maltais*. Wiesbaden: Harrassowitz.
- Rossi, E. 1936. "Malta." *Encyclopedia of Islam*, vol. 3, pp. 213–14.
- Saada, Lucienne. 1986. "Maltais en Tunisie." *Gli interscambi culturali e socio-economici fra l'Africa settentrionale e l'Europa mediterranea: Atti del congresso internazionale di Amalfi, 5-8 dicembre 1983*. Naples: n.p.

- Schat, P. 1976. *Laut- und Formenlehre des Maltesischen anhand zweier Mundarten*. Erlangen: Palm & Enke.
- . 1980. "Text aus Malta." In Fischer and Jastrow 1980: 286–91.
- Sirat, A., and E. E. Knudsen. 1973. "Notes on the Arabic Dialect Spoken in the Balkh Region of Afghanistan." *Acta Orientalia* 35: 99–101.
- Stetkevich, Jaroslav. 1970. *The Modern Arabic Literary Language: Lexical and Stylistic Developments*. Chicago: University of Chicago Press.
- Stumme, Hans. 1904. *Maltesische Studien: Eine Sammlung prosaischer und poetischer Texte in maltesischer Sprache nebst Erläuterungen*. Leipzig: Hinrichs.
- Vadala, G. 1906. *Malte et ses dépendances*. Malta: n.p.
- Wettinger, Godfrey. 1985. *The Jews of Malta in the Late Middle Ages*. Malta: Midsea Books.
- . 1986. "The Arabs in Malta." In *Malta: Studies of its Heritage and History*. Malta: Mid-Med Bank.
- , and M. Fsadni. 1968. *Peter Caxaro's Cantilena*. Malta: n.p.
- . 1983. *L-Għanja ta' Pietru Caxaru* [Peter Caxaro's Cantilena], 2nd ed. Malta: n.p.

This page intentionally left blank.

Israeli Hebrew Phonology

Shmuel Bolozky

University of Massachusetts, Amherst

17.1. Introduction

Whether or not it is appropriate to talk of the “revival” (see Wexler 1990 and references to earlier discussions) of Hebrew as a verbal means of communication in Palestine (now Israel) about a century ago, the phonological inventory of Modern Hebrew as spoken in Israel today is significantly different from that of either Biblical Hebrew (henceforth BH; see Chapter 6) or Mishnaic Hebrew. Today’s “Arabicized” Hebrew (see Blanc 1964 and elsewhere) maintains a few more phonemic distinctions than does Standard (or General) Israeli Hebrew (henceforth Modern Hebrew, MH)—notably the preservation of pharyngeals ‘ and *ħ*, and for some groups the uvular stop *q*, the velarized denti-alveolar stop *ɮ*, and the glide *w*. It also tends to preserve a rolled denti-alveolar *ɹ* and here and there a velarized denti-alveolar *ʃ*. This chapter, however, focuses on MH phonology. The “compromise” between the Sephardi and Ashkenazi traditions of pronunciations of Hebrew that constitutes MH is a considerably simpler inventory of phonemes:

Table 17-1. Israeli Hebrew Consonants

Consonant	Bilabial	Labio-Dental	Denti-Alveolar	Palato-Alveolar	Palatal	Velar	Uvular	Glottal
Stop	p b		t d			k g		ʔ
Fricative		f v	s z	ʃ (ž)			x	h
Affricate			c	(č) (j)				
Nasal	m		n					
Liquid			l				r	
Glide					y			

The consonants in brackets occur only in borrowings. Comparison with BH suggests the following “mergers”/simplifications and other changes/correspondences:

- (a) *t* and *ṭ* > *t*
- (b) *k* and *q* > *k*
- (c) ' and ' (and for some speakers *h* as well) > ' or $\emptyset$ (Optional realization as ' is most likely as onglide to a heavily stressed vowel.)
- (d) *h* > $\emptyset$ (Optional realization as *h* is most likely as onglide to a heavily stressed vowel.)
- (e) *s* and *ś* > *s*
- (f) *ḥ* and *x* as an allophone of /*k*/ > *x*
- (g) *ṣ* > *c*
- (h) denti-alveolar roll *r* > uvular *r*
- (i) *w* and *v* as an allophone of /*b*/ > *v*
- (j) MH no longer maintains gemination, except phonetically, across morpheme boundary: *yašān+nu* ‘we slept’, *avāt+ti* ‘I worked’ (~*avádeti*).
- (k) The stop–fricative alternation of BH is maintained only for *p*, *b*, and *k*, and even that part of the rule is rife with exceptions, making it very opaque. Moreover, segments that were historically geminate are not subject to spirantization, nor is historical /*q*/. *v* from BH *w* and *x* from BH *ḥ* add to further surface opacity. None of the fricatives concerned is an allophone of the corresponding stop.

Vowels: *i e a o u*

Changes from the BH Tiberian inventory:

- (a) *e*, *ε*, and *ə* > *e* (phonetically *ε*)
- (b) *a* and *ɔ* > *a*

It should be pointed out that the gap between the phonemic system of MH and earlier phases of the language cannot be regarded as evidence for MH having lost its Semitic character; phonological, phonetic, or syntactic deviations from what is conceived of as a language family norm quite commonly result from the inherent instability of these components of the grammar and should not be interpreted as evidence for severed relationship with the common source.

17.1.1. *Characterizing the phonetic–phonological–morphological continuum*

The following discussion will present some essential features of MH phonology in a phonetic–phonological–morphological continuum of processes, which can be motivated by a number of criteria. The primary criterion is universal **phonetic/phonological naturalness**, which declines from one end of the continuum to the other. Another is the degree of **automaticity** of a process from a **phonetic/phonological** point of view. An automatic phonetic process applies whenever its structural description is met. The majority of phonetic and phonological rules, however, are restricted in one way or another to particular categories (mostly morphological), to segments of the lexicon, to individual items, etc., and have lexically marked exceptions. It is also possible to view automaticity from a different angle: the degree to which a speaker is **aware** of the process having taken place. Normally, speakers are not conscious of the application of automatic phonetic rules, whereas the more morphologized a phonological process is, the more likely is the speaker to be aware of its existence—although there is still the question of whether s/he is aware of an actual process, or merely of the relationship between two related morphological patterns, one of which is our assumed process “output.” A fourth criterion, which characterizes the continuum only partially, is the extent to which a process is **optional**, or more accurately **variable**, depending on various sociolinguistic factors (age, gender, education, socio-economic status, speech style or tempo, etc.). Owing to limitations of space, only a few illustrations will be provided for each of the categories below. For additional survey-type data, see Bolozky (1978).

17.2. Natural, automatic phonetic processes

17.2.1. *‘Mechanical’ secondary stress*

In Modern Hebrew, as well as in Polish and in other languages, there exists almost-automatic alternation between stressed and unstressed syllables, where the main stress constitutes the base, and in most cases every other syllable before it receives secondary stress. Occasionally, **two** unstressed syllables are squeezed between two stressed ones. The existence of two consecutive unstressed syllables generally results from the non-realization of secondary stress when a penultimately stressed word is followed by a bi-syllabic word with final stress, as in *dóar avír* ‘air mail’, or when the secondary stress—or the lexical stress of grammatical words—would have

“clashed” (see Bolozky 1982) with the main stress of adjacent words, as in *xàmišá mìxtavím* ‘five letters’ > *xamíšá mìxtavím*, *ù kará lànú séfer* ‘he read us a book’ > *ù kará lanu séfer*. The main stress of most grammatical words is essentially equal to that of regular secondary stress, and is thus subject to destressing in the same way. It is also more flexible than the primary stress of regular lexical items. Unlike the BH situation, where stress retraction (*nəsoɡ ʾəhor*) occurred even in lexical items, as in Gen 1:5:

vəlahóšex qəró láylə
and (to) the dark he called night

> *vəlahóšex qóró láylə*

MH does allow adjacent main stresses across lexical items, as in *vədarás kélev*:

ù naág bemèirút mufrézet vədarás kélev
he drove in speed excessive and ran over a dog

In grammatical words, on the other hand, clash of the main stress with another main stress results in reconfiguration of the stress pattern in a manner that avoids such clash:

atà bá ‘are you coming?’
you come

> *àta bá*

ù kará lànú abáyta ‘he called us home’
he called to us home

> *ù kará lanù abáyta*

Under relatively rare conditions of very low contextual prominence (e.g. next to contrastive stress), the main stress of some lexical items is low, equivalent to that of secondary stress, and thus may undergo similar reconfiguration (see Kadmon 1983):

ù kará pérek xaméšesrè lo šéšesrè
he read chapter fifteen not sixteen

> *ù kará perèk xaméšesrè lo šéšesrè*

The opposition between stressed and unstressed syllables is a common phenomenon, and fixed repetition of this contrast in the form of rhythmic alternation is a natural phenomenon in the colloquial registers of a number of languages. Regular rhythmic alternation is frequently found in folk poetry, and is also characteristic of counting sequences or memorized numeral sequences. As shown in Bolozky and Haydar (1986), preference for the unsuffixed numeral set in colloquial Hebrew and for the suffixed one in the dialects of Arabic (e.g. Lebanese Arabic below) originates, at least in part, from

numeral paradigms in trochaic rhythm as chanted and acquired for the first time by children:

MH: (a)xát—štáym šalóš arbá xaméš—šésš—šéva šmóne téša éser

Arab: wáḥed tnéyn—tléte 'arb'a xámse sítte sáb'a tméni tís'a 'ášra

Had the situation been reversed, the chanted sequences would not have flowed as naturally (a partial anapest in Hebrew, and staccato in Lebanese Arabic):

exád šnáim šlošá arbaá xamišá šišá šivá šmoná tišá asará

... tlét 'arb' xáms sítt sáb' tmén tís' 'ášr

Alternating stress can also be observed in children's rhymes, e.g.

mìmmromím pcacá yorédet ...

from the sky bomb comes down

and in rhythmic counting by children for 'random selection' purposes, as in

éven nyár umìsparáim ...

rock paper and scissors

én den díno sóf a lá katíno ... (random selection rhyme of unclear origin, probably Mediterranean Lingua Franca)

as well as in rhythmic chants of groups at sports events and such (see Gil 1986):

él—él—ìsraél (sports chant to encourage Israeli team)

àšofét abáyta 'referee go home!'

the referee home

Mechanical secondary stress assignment is a natural process, then, applying automatically whenever its structural description is met, and speakers are certainly unaware of its having taken place. It is also variable, at least in that it "cancels out" when clashing with an adjacent primary stress.

17.2.2. Voicing assimilation

MH voicing assimilation of obstruents is anticipatory/regressive, which is the unmarked direction for this kind of assimilation (progressive assimilation is unmarked when stems determine the realization of the suffix. Unless marked otherwise, the main stress always falls on the final vowel):

No Contact	Formal Alternant	Informal Variant
sagar 'he closed'	yisgor 'he will close'	yizgor
pagaš 'he met'	pgiša 'meeting'	bgiša
zaken 'old (m.sg.)'	zkenim 'old (m.pl.)'	skenim
dakar 'he stabbed'	yidkor 'he will stab'	yitkor

As shown in Bolozky (1978, 1985), there appear to be two types of exceptions: normally, *v* does not cause voicing assimilation, and *x* does not undergo it:

<i>kvar</i>	‘already’	<i>exzir</i>	‘returned (trans.)’
<i>kviš</i>	‘road’	<i>yixboš</i>	‘will conquer’
<i>tikva</i>	‘hope’	<i>exdir</i>	‘caused to penetrate’
<i>tvia</i>	‘claim; drowning’		
<i>itva</i>	‘outlined’		

It appears that *v* not causing voicing assimilation is due to substratum influence of Slavic languages such as Russian and Polish, where (owing to *v* originating from a historical glide *w*) the same phenomenon is observed (and furthermore, where at least intra-morphemically in Warsaw Polish, **progressive** assimilation occurs instead—which may also surface in the speech of some Polish-born Israelis). *x* does not readily undergo voicing assimilation apparently because speakers might be reluctant to assimilate *x* into something that sounds too much like an *r* that should not be there (Hebrew does not have a truly corresponding voiced fricative counterpart *ɣ*: *ʁ*, the nearest voiced segment, does not have enough friction to qualify as an obstruent.) In (very) casual speech, however, these exceptions tend to be eliminated; the more casual the register and more rapid the tempo, the greater the likelihood of the process having no exceptions:

<i>gvar</i>	‘already’	<i>eyzir</i>	‘returned (trans.)’
<i>gviš</i>	‘road’	<i>yiyboš</i>	‘will conquer’
<i>tigva</i>	‘hope’	<i>eydir</i>	‘caused to penetrate’
<i>dvia</i>	‘claim; drowning’		
<i>idva</i>	‘outlined’		

We are dealing, then, with a natural phonetic process (uniformity of voicing clearly facilitates production), working in the unmarked (anticipatory/regressive) direction, which applies automatically whenever the structural description is met, and exceptions tend to be eliminated in fast/casual speech. It is normally not a conscious process; speakers will deny, for instance, that they ever pronounce *yisgor* as *yizgor*, etc. It is also variable, relative to the degree of casualness and tempo involved. The more casual the register and the faster the tempo, the more likely the assimilation.

17.3. Natural phonological processes with some restrictions

17.3.1. Avoidance of identical or homorganic consonant sequences

MH does not maintain the gemination of BH, but geminates may be formed across a morpheme boundary:

<i>dan</i>	'he discussed'	<i>dánnu</i>	'we discussed'
<i>šavat</i>	'he was on strike'	<i>šavátti</i>	'I was on strike'
<i>tamim</i>	'naïve'	<i>hittamem</i>	'he feigned naïveté'

Within the stem, however, geminates are either broken with the minimal vowel *e* (the assumption that this *e* is epenthetic is based on pattern comparison):

	<i>zalelan</i>	'glutton'
	<i>xatetan</i>	'meddler'
cf.	<i>kamcan</i>	'miser'
	<i>noxexut</i>	'presence'
	<i>holelut</i>	'folly, hilarity'
cf.	<i>roxxut</i>	'pharmacology'

or the elision of a vowel separating between identical consonants is blocked (although reduction to *e* still takes place), as in

	<i>xagag</i> 'he celebrated' ~ <i>xagega</i> 'she celebrated'
cf.	<i>katav</i> 'he wrote' ~ <i>katva</i> 'she wrote'
	<i>kucac</i> 'it was cut' ~ <i>kucecu</i> 'they were cut'
cf.	<i>šupac</i> 'it was overhauled' ~ <i>šupcu</i> 'they were overhauled'
	<i>itpalel</i> 'he prayed' ~ <i>itpalela</i> 'she prayed'
cf.	<i>itlabeš</i> 'he got dressed' ~ <i>itlabša</i> 'she got dressed'

except for very casual/fast speech, in which elision may marginally occur, particularly when fricatives are involved (e.g. *em šâxexú* 'they forgot' > *èm šâxxú*). McCarthy (1986) regards the blocking of *e*-deletion in *xagega*, etc. as "antigemination": syncope rules are prohibited from creating clusters of identical consonants. This is an immediate corollary of his Obligatory Contour Principle (OCP), which prohibits adjacent identical elements at the melodic level (either consonantal or vocalic, in an autosegmental analysis). The OCP does not apply to the across-morpheme-boundary cases, since different morphemes are represented on different tiers.

The *šavátti* type above often undergoes *e*-insertion, which breaks the surface geminate or a homorganic *d+t* sequence (just as in English *prodded*, *wanted*—except that in English the process is obligatory):

<i>šavátti</i>	'I was on strike' ~ <i>šaváteti</i>
/avád+ti/	'I worked' > <i>avátti</i> ~ <i>avádeti</i>

and needless to say, voicing assimilation and the splitting process will have to be mutually exclusive, to avoid **aváteti* (which can be achieved by Kiparsky's 1973 Elsewhere Condition, according to which the more specific rule applies before the more general one). Its not applying to *hittamem* suggests that the process is restricted to sequences involving an **inflectional** affix, and it is blocked in *dánnu* since **danenu* would have been interpreted as stemming from a geminate root *d.n.n* instead of the correct *d.w.n*.

The avoidance or splitting of identical or homorganic consonants is phonetically natural, but nevertheless restricted; speakers may or may not be aware of its having taken place.

17.3.2. *The new imperative in colloquial Hebrew*

In Bolozky (1979) it was argued that in colloquial Hebrew, commands are normally realized in future-tense form used imperatively. What appears to be partial colloquial resurrection of the normative formal imperative, as in *štok* 'shut up!', *lex* 'go!', etc., is in fact a future (BH imperfect) form whose prefix has been "chopped off":

Fut/Imp.	Reduction 1	Reduction 2	
<i>tešev</i>	<i>tšev ~ čev</i>	<i>šev</i>	'sit down! (m.sg.)'
<i>telamed</i>	<i>tlamed</i>	<i>lamed</i>	'teach!'
<i>titlabeš</i>		<i>tlabeš</i>	'get dressed!'
<i>tizaer</i>	<i>tzaer > đzaer</i>	<i>zaer</i>	'watch out!'

That this is **not** the normative imperative can be proven by the absence of the prefixal *h* required for the formal imperative in Hitpa'el and Nif'al (i.e. *hitlabeš* and *hiza(h)er*, respectively, would have been expected), as well as by the clearly-reduced new **suffixed** imperatives in Pa'al, which are distinct from their normative counterparts (illustrations are with the feminine suffix +*i*⁵; the same applies to the plural suffix +*u*⁵):

Normative	Fut/Imp.	Reduction 1	Reduction 2
<i>sigri</i>	<i>tisgeri</i>	<i>tsgeri ~ cgeri ~ đzgeri</i>	<i>s/zgeri</i> 'close!'
<i>pitxi</i>	<i>tiftexi</i>		<i>ftexi</i> 'open!'

This is a natural phenomenon; many languages allow imperative use of their future forms (see Ultan 1978). Shortening increases the stress and urgency of the command: in BH, the shorter jussive form was used for negative imperatives (*ʾal tešt* 'do not drink!', full form *tištē*) and for command or request in the third person (*yəhi ʾor* 'let there be light!'). Although the first person "command," the cohortative, actually involves lengthening (*ʿšməra* 'I will observe, let me observe' vs. imperfect *ʿšmor*), it does not involve the urgency

of a shortened form, and the same can be said of other long-form imperatives (*qúma* ‘arise’ is less urgent than its variant *qum*). The process is also natural in that only deletion of the “minimal” vowels *e* and *i* is allowed; the *a* of Hiʿil (*tasbir* ‘explain!’) is never affected. It is not automatic, though; it is essentially restricted to the future-used-imperatively, and only occasionally expands to non-imperative future forms. Speakers are at least somewhat aware of its having taken place, as can be shown from occasional orthographic evidence. For a recent, most detailed description of MH imperatives and their substitutes, see Henkin (to appear).

17.4. Morphologically constrained phonological processes that have no exceptions, but are not (or are no longer) phonetically motivated

17.4.1. Residues of Philippi’s Law

Essentially, the general lowering of a stressed *i* or *e* in a closed syllable to *a* is now restricted to the verb system:

<i>diber</i>	‘he spoke’	~ <i>dibárti</i>	‘I spoke’
<i>isbir</i>	‘he explained’	~ <i>isbárnu</i>	‘we explained’
<i>itlabeš</i>	‘he got dressed’	~ <i>itlabášta</i>	‘you got dressed’

In the verb, however, it has no exceptions: *šévna* ‘sit! (f.pl.)’ etc. no longer exist in MH. It is not a variable rule, and speakers are aware of the alternation. For additional discussion, see Bolozky (1978).

17.4.2. Hitpaʿel metathesis

In Hitpaʿel, and in related derived nominalizations (e.g. *istalkut* ‘going away’, cf. *istalek* ‘went away’), a coronal sibilant metathesizes with the prefixal *t*:

	Underlying	Surface	
(a)	/hit+labeš/	<i>itlabeš</i>	‘got dressed’
(b)	/hit+saken/	<i>istaken</i>	‘took risk’
	/hit+šameš/	<i>ištameš</i>	‘used’
	/hit+zaken/	<i>izdaken</i>	‘became old’
	/hit+camek/	<i>ictamek</i>	‘shrank’

When voicing assimilation is involved (see § 17.2.2), the two processes should apply simultaneously, to block **istaken* for ‘he became old’ (see Bolozky 1978). Ordering voicing assimilation before metathesis would have also worked, but one would not wish to order a phonetic process before a morphologically restricted rule like this one.

Since this is an historical, established process, it might make more sense to ask not what the actual phonetic motivation for this rule is, but rather why it is not inverted or lost. One possible explanation is avoidance of single-unit affricate interpretation of a *t*-plus-sibilant, which would make the Hitpa‘el prefix opaque, causing the forms concerned to look like infinitives or imperatives of Nif‘al (e.g. /hit+saken/ > **hitsaken* > **hicaken*). It would also make the underlying root opaque. So the process can be explained on phonological-semantic grounds, and has no exceptions among Hitpa‘el forms, but can hardly be motivated phonetically: even Hif‘il forms with the same structural description, such as *hitis* ‘caused to ferment’, *hitšiš* ‘tired out (trans.)’, do not undergo the rule, let alone other items (/t+šuv+a/ ‘answer’ > *tšuva*, not **štuva*). It is strictly restricted to Hitpa‘el. It is not a variable process; determining speakers’ awareness of its application would require testing.

17.5. Other morphologically constrained phonological processes that are not (or are no longer) phonetically motivated

17.5.1. Pretonic and antepretonic a/e-deletion

In the verb, a non-high vowel is elided in a pretonic open syllable:

<i>katav</i>	‘he wrote’	~ <i>katva</i>	‘she wrote’
<i>yikanēs</i>	‘he will enter’	~ <i>yikansu</i>	‘they will enter’
<i>diber</i>	‘he spoke’	~ <i>dibru</i>	‘they spoke’
<i>sudar</i>	‘he/it was arranged’	~ <i>sudra</i>	‘she/it was arranged’
<i>itlabeš</i>	‘he got dressed’	~ <i>itlabšu</i>	‘they got dressed’

Exception: the present/present participle of Pu‘al (*mesudarim* ‘are (m.) arranged’, not **mesudrim*). If a sequence of (any) three consonants or two identical ones (see above) would be formed in the process, they are broken/avoided by an intervening e:

<i>tixtov</i>	‘you (m.) will write’	~ <i>tixtevi</i>	‘you (f.sg.) id.’
<i>nixtav</i>	‘it was written’	~ <i>nixtevu</i>	‘they were written’
<i>uxtav</i>	‘it was dictated’	~ <i>uxteva</i>	‘she/it was dictated’

Two exceptions: the present/present participle of Nif‘al (*nixnasim* ‘enter (m.pl.)’, not **nixnesim*) and of Huf‘al (*mustarim* ‘are (m.) hidden’, not **mustirim*).

This **could** have been a natural phonetic rule, since pretonic reduction and elision are (universally) expected. It is, however, restricted to the verb system, and even there one finds exceptions that are hard to account for. Reduction/elision does apply to one non-verbal pattern, as in:

M.Sg.	F.Sg.	M.Pl.	F.Pl.	
<i>tipeš</i>	<i>tipša</i>	<i>tipšim</i>	<i>tipšot</i>	‘fool(ish)’
<i>xiver</i>	<i>xivéret</i>	<i>xivrim</i>	<i>xivrot</i>	‘pale’

and to a few other items, like:

<i>totax</i>	‘gun’	~ <i>totxan</i>	‘gunner’
<i>mišpat</i>	‘law’	~ <i>mišpetan</i>	‘jurist’

but in most cases it is blocked because antepretonic *a*-deletion in non-verbal categories takes precedence:

M.Sg.	F.Sg.	M.Pl.	F.Pl.	
<i>šamen</i>	<i>šmena</i>	<i>šmenim</i>	<i>šmenot</i>	‘fat’
<i>šafan</i>	<i>šfana</i>	<i>šfanim</i>	<i>šfanot</i>	‘rabbit’
<i>davar</i>		<i>dvarim</i>		‘thing’

Antepretonic *a*-deletion is also heavily morphologized, and is certainly less motivated phonetically. Historically, reduction of either pretonic or antepretonic vowels appears to have applied to any unstressed non-high vowel in an open syllable, regardless of distance from the main stress, that was not affected by pretonic lengthening or other tensing processes. Thus, since the second *a* in /davar+im/ was lengthened, and the second one in /šafann+im/ was closed by gemination, reduction could only apply to the first *a*. In MH, though, with no lengthening and no gemination, there is no particular phonetic reason for deleting a vowel **two** syllables before the stress, and the restriction to *a* to the exclusion of *e* (*terucim* ‘excuses’ > **trucim*, *pexamim* ‘coals’ > **pxamim*, etc.—see Bolozky and Schwarzwald 1990) makes it even less plausible as a phonetically motivated process. Another point: if complete deletion cannot take place, at least *a* is reduced to *e* (*yašar* ‘straight’ ~ *yešarim* (pl.), *ša(‘)ul* ‘borrowed’ ~ *šeulim* (pl.)). But reduction does not apply with *a* that was preceded by a now-lost low consonant:

<i>ašir</i>	‘rich, m.sg.’	~ <i>aširim</i>	‘rich, m.pl.’, not * <i>eširim</i>
<i>avud</i>	‘lost (m.sg.)’	~ <i>avuda</i>	‘lost (f.sg.)’
<i>arug</i>	‘killed (m.sg.)’	~ <i>arugim</i>	‘killed (m.pl.)’

Historically, low consonants preferred a low vowel. Today, however, with no phonetic realization of such consonants, the blocking of reduction is hard to motivate. Furthermore, the loss of gemination in MH has also removed the motivation for blocking *a*-deletion in forms like:

Singular	Historical	Plural	Historical	
<i>nagar</i>	<i>naggar</i>	<i>nagarim</i>	<i>naggarim</i>	‘carpenter’
<i>paṭiš</i>	<i>paṭtiš</i>	<i>paṭišim</i>	<i>paṭtišim</i>	‘hammer’
<i>šapud</i>	<i>šappud</i>	<i>šapudim</i>	<i>šappudim</i>	‘skewer’

Historically, deletion was blocked because its application would have created an impermissible three-consonant cluster, a geminate counting as two consonants. With MH degemination, *a*-deletion **should** have applied, had it been a phonetically motivated process.

Both deletion/reduction processes, then, have lost their phonetic motivation—*a*-deletion perhaps more so than *e*-deletion—and are morphologically restricted. They also are not variable in MH (*yoševim* for *yošvim* ‘sit (m.pl.)’, etc. may rarely be heard among former speakers of Judeo-Spanish—a very marginal phenomenon). Since these are not phonetic processes, it is hard to tell whether speakers actually derive one form from the other by deleting a vowel (e.g. *katav* ‘he wrote’ ~ /katav+a/ ‘she wrote’ > *katva*), or simply make the connection between two related patterns—e.g. CaCaC and a related CaCC+a feminine pattern. (Speakers are clearly aware of the relationships between patterns.)

17.5.2. Assignment of main stress

MH main stress is normally assigned to the final syllable, with a number of classes of exceptions that are sufficiently well defined for systematic treatment. One can first take care of these exceptions, and then let all the remaining forms, which constitute the majority, be assigned the unmarked final stress by the universal Elsewhere Condition.

In the verb system, the final vowel of the stem and main stress assignment appear to be mutually exclusive (see Bolozky 1978): a stem-final vowel that has not been affected by elision or reduction is assigned main stress. As noted above, pretonic deletion in the verb can hardly be regarded as a phonetically motivated process in MH. If so, one could let deletion or reduction of a non-high vowel at the end of the verb stem be triggered by a +V# or +VC# suffix irrespective of whether it is stressed or not, mark a few exceptions (*egena* ‘she protected’, *exela* ‘she began’), and subsequently assign primary stress to all remaining final stem vowels (excluding derived ones resulting from reduction):

<i>katávti</i>	‘I wrote’	<i>dibárta</i>	‘you spoke’
<i>káma</i>	‘she rose’	<i>hitlabášti</i>	‘I got dressed’
<i>kanínu</i>	‘we bought’	<i>isbíru</i>	‘they explained’
<i>macáti</i>	‘I found’	<i>egéna</i>	‘she protected’
<i>yakúmu</i>	‘they will rise’	<i>ekímu</i>	‘they established’
<i>nišbárti</i>	‘I broke’	<i>ipíla</i>	‘she dropped (trans.)’
<i>nimcénu</i>	‘we were found’		

Present/present participle forms will have to be excluded, to block **mazkíra* ‘remind (f.sg.); secretary’, **mešubášim* ‘distorted (m.pl.), faulty’, **mušlámim* ‘completed (m.pl., perfect)’—which is, perhaps, one of the arguments for considering the present participle as part of the non-verbal system.

As for the non-verbal system, there are a few sporadic exceptions, such as *láma* ‘where’, (‘)éfo ‘where’, (h)éna ‘here’, but in most cases penultimate stress applies to reasonably well defined groups. The main exception to final stress is the segolate class. Historically, this was hardly a problem, since stress assignment applied **before** the epenthesis of *ε* to break an unpermitted final cluster, when the surface penultimate vowel was still the final vowel:

/malk/ ‘king’ > *málk* > *málek* > *mélek* > *mélex*

Assuming an underlying /malk/ was based on the existence of alternants in which this stem surfaces, e.g. *malka* ‘queen’. As pointed out in Boložky (1978), however, in MH most segolates do **not** have CVCC alternants (if they do, they are confined to the literary register), e.g.

<i>téred</i>	‘spinach’	<i>xóken</i>	‘enema’
<i>téfer</i>	‘seam’	<i>šéten</i>	‘urine’
<i>rótev</i>	‘sauce’	<i>bóreg</i>	‘screw’
<i>gézer</i>	‘carrot’	<i>dófek</i>	‘pulse’
<i>délek</i>	‘fuel’		

which suggests that segolate stress be assigned by surface sequences, i.e. to the penult in ...CeCeC# and ...CoCeC# nouns that are not directly derived from verbs (to exclude verbs, e.g. *kotev* ‘write (m.sg.)’, *berex* ‘he blessed’—as well as verb-related agent nouns like *šoter* ‘policeman’). If that final eC# sequence is an +et^f suffix, verbs are penultimately stressed as well (i.e. not only *miktéret* ‘pipe’, but also *kotévet* ‘write (f.sg.)’).

When consequences of formerly low consonants are included, the list may also include ...Ca(x)aC#:

<i>ráaš</i>	‘noise’	<i>mikláxat</i>	‘shower’
<i>náxal</i>	‘river’	<i>šomáat</i>	‘hear’
<i>mišmáat</i>	‘discipline’	<i>šoláxat</i>	‘send (f.sg.)’

CeCe^f (*péle* ‘miracle’, *déše* ‘lawn’), CeCa(x)^f (*téva* ‘nature’, *kérax* ‘ice’), Co(x)aC^f (*nóar* ‘youth’, *dóar* ‘mail’, *šóxad* ‘bribe’). Exceptions like *nahar* ‘river’, *naxaš* ‘snake’, *šena* ‘sleep’, *koxav* ‘star’, *ge’e* ‘proud’, *kehe* ‘dark’, which are a minority, will be marked as such.

Another group of words that is stressed penultimately involves a “furtive *patax*”—an *a* inserted before historical word-final *h*, ‘, and *ḥ* when those were preceded by a vowel-other-than-*a*:

<i>gavóa</i>	‘tall (m.sg.)’	historically <i>gavóah</i>
<i>šoméa</i>	‘hear (m.sg.)’	historically <i>šoméa’</i>
<i>potéax</i>	‘open (m.sg.)’	historically <i>potéah</i>

If underlying low consonants are assumed, regular final stress applies before *a*-insertion:

/gavoh/ ‘tall (m.sg.)’ > *gavóh* > *gavóah* > *gavóa*

but since the low consonant is not ever realized, a more concrete approach would be preferable: stress a vowel-other-than-*a* that is immediately followed by word-final *a* that is **not** suffixal (to block **meví+a* for *mevia* ‘bring (f.sg.)’, **bí+a* for *bia* ‘coming’. Exceptions, like *kia* from *kiha* ‘he scolded’, would be rare), or by *ax*⁵.

Bolozky (1978) also discusses the influence of Yiddish stress, which in the colloquial shifts final stress to penultimate position in proper names,

yael > *yáel*
menaxem > *menáxem*
naftali > *naftáli*

occasionally creating “minimal pairs” of common and proper nouns,

rexovot ‘streets’ vs. *rexóvot* ‘Rehovot’
xaim ‘life’ vs. *xáim* ‘Chaim’

or of Hebrew and related Yiddishized versions:

taxlit ‘aim, purpose’ vs. *táxles* ‘business’
meci(‘)a ‘finding, bargain’ vs. *mecíe* ‘bargain (usu. ironic)’

as well as the colloquial tendency to shift stress to the first syllable when the word concerned is used in the context of games:

klafim ‘cards’ vs. *kláfim* ‘card game’
bulim ‘stamps’ vs. *búlim* ‘(collecting) stamps’
rišon ‘first’ vs. *ríšon* ‘first step in game’
xamiši ‘fifth’ vs. *xámiši* ‘fifth step in game’
monopol ‘monopoly’ vs. *mónopol* ‘Monopoly (game)’

This phenomenon seems to “spill over” to some common colloquial variants, where the alternative penultimate stress stays on the same syllable even after suffixation:

Singular		Plural		
Formal	Non-formal	Formal	Non-formal	
<i>tiras</i>	<i>tíras</i>	<i>tirasim</i>	<i>tírasim</i>	‘corn’
<i>pilpel</i>	<i>pílpel</i>	<i>pilpelim</i>	<i>pílpelim</i>	‘pepper’

In borrowed non-verbs, stress normally (but not always) maintains the position it holds in the language of origin, and stays in a steady relationship to

specific suffixes (see Bolozky 1978). The addition of native suffixes does not affect the position of stress in borrowed words. Some familiar native words demonstrate similar behavior in all non-formal registers, not only when the basic stress is penultimate, as in *tíras* ‘corn’ above, but when it is final as well (see Bat-El 1989):

Singular	Plural		
Formal	Formal	Non-formal	
<i>marak</i>	<i>merakim</i>	<i>marákim</i>	‘soup’
<i>sabon</i>	<i>sabonim</i>	<i>sabónim</i>	‘nerd’

Note, however, that in its basic meaning, the plural of *sabon* ‘soap’, *sabonim*, has regular final stress.

It appears, then, that main word-stress can be reasonably well defined for a variety of penultimate (and some word-initial) environments, and once these are taken care of, final word stress applies elsewhere. As for speakers’ awareness of it—one indication that they do is that they distinguish different meanings for minimal pairs that differ only in the location of their main stress.

17.5.3. *Residues of historical low consonants*

With the loss of low consonants or their mergers with other segments in MH, their role in the phonological component requires reevaluation. In the case of *ħ*, for instance, the merger with *x* has indeed resulted in some degree of regularization, particularly in dispensing with the need for *a*-epenthesis to avoid a syllable-final *ħ*:

<i>sixka</i>	‘she played’ (rarely <i>sixaka</i>)
<i>maxku</i>	‘they erased’ (rarely <i>maxaku</i>)
<i>saxkan</i>	‘actor’ (<i>saxakan</i> only in very formal register)

but also in allowing variable regularization of the prefix, “returning” it to *i*, as in

<i>exzik</i>	‘he held’	~ <i>ixzik</i> (cf. <i>isbir</i> ‘he decided’)
<i>exlit</i>	‘he decided’	~ <i>ixlit</i>

and in variable choice of the stem-vowel in the future of Pa‘al, with an option for using the regular *o* alongside the “guttural”-related *a* when *x* from *ħ* is the **second** radical of the root:

<i>yivxar</i>	‘he will choose/elect’	~ <i>yivxor</i>
<i>yivxan</i>	‘he will examine’	~ <i>yivxon</i>

The “ex-guttural” impact is still strongly felt, however, in word-final position: choice of *a* rather than *o* in the future of Pa‘al:

<i>yīšlax</i>	‘he will send’	~ * <i>yīšlóax</i>
<i>yivrax</i>	‘he will flee’	~ * <i>yivróax</i>

and insertion of *a* if the preceding vowel is not *a*:

/šalix/	‘messenger’	> <i>šalíax</i>
/samex/	‘happy’	> <i>saméax</i>

It is also maintained in the replacement of regular ...*éxet*⁵ by ...*áxat*⁶:

<i>šoláxat</i>	‘send (f.sg.)’ (cf. <i>kotévet</i>)
<i>mefatáxat</i>	‘develop (f.sg.)’ (cf. <i>medabéret</i>)

Since in most cases *x* that corresponds to the historical allophone of *k* can **also** occur in the same environments (except for the *yivxar* group), *x* from *ḥ* still requires marking in some way to account for these consequences.

As shown in detail in Bolozky (1978), the ex-gutturals or their traces fulfill a role **even if not realized**: as occupiers of consonantal slots, to give speakers clues as to *miškal* (canonical morphological pattern) membership, as well as to account for deviation from regular *miškal* patterns, owing to processes which were historically natural, but are no longer transparent. There are essentially two ways of accounting for speakers’ capability to relate deviant forms with formerly-low consonants to their regular *miškal*-base. One is to start from abstract representations based on regular *miškalim*, derive the forms concerned by means of morpho-phonological processes, then dispose of whatever is not realized phonetically. Assuming underlying low consonants would “motivate” the need to get rid of a low consonant at the syllable coda:

/maca/	‘he found’	> <i>maca</i>
/maca+ti/	‘I found’	> <i>macáti</i>
/gavah/	‘he was tall’	> <i>gava</i> (MH only)
/šama/	‘he heard’	> <i>šama</i> (MH only)

or to add a vowel, which would shift them to syllabic onset position:

/šo‘el+im/	‘ask (m.pl.)’	> <i>šoʔl+im</i> > <i>šoʔalim</i> > <i>šoalim</i>
(cf. /kotev+im/	‘write (m.pl.)’	> <i>kotvim</i>)

and if a prefix is involved, it also echoes that vowel across it:

/ti+ʔbod/	‘you will work’	> <i>taʔabod</i> > <i>taavod</i>
(cf. <i>ti+xtov</i>	‘you will write’)	
/hi+ʔmin/	‘he believed’	> <i>heʔemin</i> > <i>eemin</i>
(cf. <i>isbir</i>	‘he explained’)	
/mi+ʔrav/	‘west’	> <i>maʔarav</i> > <i>maarav</i>
(cf. <i>mi+zrax</i>	‘east’)	

A low consonant is also not allowed with another consonant in the syllable onset:

- /'ašir+im/ 'rich (m.pl.)' > širim > 'aširim > aširim
 (cf. /zariz+im/ 'nimble (m.pl.)' > zririm)
 /hlix+a/ 'walking' > halixa > alixa
 (cf. ktiv+a 'writing')
 /ka'uv+im/ 'painful (m.pl.)' > k'uvim > ke'uvim > keuvim
 (cf. /katuv+im/ 'written (m.sg.)' > ktuvim)
 /t+'uf+a/ 'aviation' > te'ufa > teufa
 (cf. t+rūm+a 'contribution')

And a word-final low consonant other-than-a-glottal-stop can be used to account for the "furtive *patax*," i.e. the insertion of *a* when historical final *h*, *ʕ*, and *h* (see § 17.5.2) are preceded by a vowel-other-than-*a*:

- /gavoh/ 'tall (m.sg.)' > gavóah > gavóa
 /yode'/ 'know (m.sg.)' > yodéa' > yodéa

This is an abstract approach, and in most cases is problematic in that it assumes underlying **low** consonants for which there is no sufficient synchronic motivation. Speakers can be argued to refer to consonantal positions where the "gutturals" used to be, but not to actual "gutturals." A more realistic way of accounting for speakers' capability to assign forms with former "gutturals" to prototypical *miškalim* is to assume that they form recognition strategies based on regularities observed in surface configurations, and that although they extrapolate from them the existence of consonantal slots, these discovery procedures have nothing to do with the feature "low." Below are some initial formulations of possible slot-discovery procedures.

- (i) Any syllable-initial vowel is an obvious indication of a lost "guttural" that used to function as its onset:

<i>amar</i>	'he said'	<i>meir</i>	'give light (m.sg.)'
<i>oved</i>	'he works'	<i>yedia</i>	'message'
<i>šual</i>	'fox'	<i>eaxzut</i>	'settlement'

Exception: *a* preceded by a stressed vowel-other-than-*a* and followed by a word-final *x* (e.g. *šalíax* 'messenger'), where this *a* is epenthetic (i.e. does not constitute part of the *miškal*).

- (ii) When the syllable-initial and immediately preceding syllable-final vowels are identical, the ex-guttural may be signaled by a long vowel in casual speech:

<i>ta+avod</i>	'you (m.sg.) will work'	> <i>ta:vod</i>
<i>e+evod</i>	'I will work'	> <i>e:vod</i>
<i>revi+i</i>	'fourth'	> <i>revi:</i>
<i>šiamum</i>	'boredom'	> <i>šiimum</i> > <i>ši:mum</i>

In **very** casual or fast speech, shortening (e.g. to *tavod*, *šimum*) is **also** possible, but is generally considered substandard; alternatively, it may be argued that a **trace** of length is **always** there, regardless of register. When the two adjacent vowels are identical and the first is prefixal, the second is epenthetic, while the first, which constitutes a component of the *miškal*, is underlyingly /i/ if it surfaces as *a* or *e* (*taavod*, *eevid*), and /u/ if it surfaces as *o* (*o+osak* ‘he was employed’).

- (iii) A word-final vowel (*a* or *e*, but theoretically it could be any) that is preceded by a **stressed** vowel in the preceding syllable signals a **following** (syllable-final) low consonant that has been lost (*téva* ‘nature’, *pére* ‘wild one, wildly’).
- (iv) When a word-final *a* that is **not** suffixal is immediately preceded by a stressed vowel-other-than-*a*, it signals a following (syllable-final) lost guttural, while the *a* itself is epenthetic (i.e. does not constitute part of the *miškal*):

<i>yadúa</i>	‘known (m.sg.)’	
<i>gavóa</i>	‘tall (m.sg.)’	vs. <i>heví+a</i> ‘she brought’
<i>higía</i>	‘he arrived’	vs. <i>higí+a</i> ‘she arrived’

So the historically low consonants are no longer low, and essentially function as occupiers of consonantal slots in order to account for *miškal* membership and deviations from *miškalim* that are fairly systematic but cannot be motivated independently on phonetic grounds. Normally, these deviations are not variable, with the notable exception of some processes related to *x* from historical *h*. Speakers are probably aware of the existence of these “empty” slots.

17.5.4. Stop spirantization

As noted in § 17.1k, the post-vocalic spirantization of BH stops is restricted in MH to *p*, *b*, and *k*, and even that part of the rule is very opaque, owing to numerous exceptions and constant flux. Opacity is further increased by degemination, since degeminated segments continue to block spirantization as if they were still geminated, and so does *k* from historical *q*. *v* from BH *w* and *x* from BH *h* add to further surface opacity. Residual spirantization has received considerable attention in the literature—see, for instance, Ben-Horin and Bolozky (1972), Barkai (1978), Bolozky (1980), Schwarzwald (1981)—most of which seems to suggest that it perhaps should no longer be regarded as a significant generalization. At best, one could point to tendencies for the rule to operate in a number of sub-environments, like word-

finally (which is, essentially, the only environment in which the rule is rarely contradicted, and even **this** holds true only for the **native** lexicon):

<i>kaf</i>	'spoon'	~ <i>kapot</i>	'spoons'
<i>dov</i>	'bear'	~ <i>dubim</i>	'bears'
<i>rax</i>	'soft (m.sg.)'	~ <i>raka</i>	'soft (f.sg.)'

or after a prefix ending in a vowel:

<i>katav</i>	'he wrote'	~ <i>yi+xtov</i>	'he will write'
<i>pitéax</i>	'he developed'	~ <i>me+fatéax</i>	'develop (m.sg.)'
<i>badak</i>	'he examined'	~ <i>ni+vdak</i>	'he was examined'
		<i>mi+xtav</i>	'letter'
		<i>ma+ftéax</i>	'key'
		<i>mi+vxan</i>	'test'

or one could simply list stop-spirant alternations as associated with particular paradigms (within the verb and elsewhere), as in Schwarzwald (1981)—and then try to account for the numerous deviations by additional principles. In Ben-Horin and Bolozky (1972), for instance, it was pointed out that there is some correlation between the degree of opacity and its impact on either input or output of spirantization—and the degree and nature of deviation from expected norms. Thus, one would expect more violation of historical *k~x* alternations because of opacity in both input (historical *q*) and output (historical *h*), with violations going both ways, whereas in the case of *b~v* increased opacity is only in the output (historical *w*), which consequently causes more output deviations (i.e. “overapplication” on the surface), and the same applies to *p~f*, owing to numerous borrowings with *f* in positions other than post-vocalic. But there are too many exceptions to this generalization—some of them due to even greater opacity since 1972. Barkai (1978) proposes that inapplication or overapplication be accounted for by avoidance of ambiguity—which again works only in some of the cases. It appears that the only way of accounting for the various numerous deviations is, as suggested in Bolozky (1980) and in Schwarzwald (1981), to attribute them to analogical leveling, directed towards already-existing opacity (which acts as a trigger, signaling that “it **can** be done”), or towards the unmarked form, or (less frequently) towards the realization that minimizes opacity. Furthermore, with some notable exceptions, the tendency for analogical leveling decreases with the decrease in derivational bond: it is stronger within inflections, but weakens with the increase in distance between an inflectional form and a related derivational one that is not automatically predictable (see Bolozky 1980). Below are a few illustrations, from Bolozky (1980).

At the end of the word (in native words) one always finds a spirant. Analogical leveling is not common; when occurring, however, the formal alternant is rarely used. Analogy is with the unmarked form with a spirant:

Formal	Alternant	Colloquial Var.
<i>ratov/uv</i> ‘wet (m.sg.)’	<i>retuba</i> (f.sg.)	<i>retuva</i>
<i>metofef</i> ‘drummer’	<i>tipuf</i> ‘drumming’	<i>tifuf</i>
<i>musax</i> ‘garage’	<i>musakim</i> (pl.)	<i>musaxim</i>

In the beginning of the stem (particularly in Pi‘el), the *k*-*x* variation undergoes analogy in the direction of both stop and spirant—probably owing to the above-mentioned opacity with both output **and** input (*k* from BH *q* and *x* from BH *ḥ*); in *b*-*v* and *p*-*f* variations analogy is usually with the spirant—because of opacity with the output (*v* from BH *w*, and *f* because of opacity caused by numerous borrowed verbs with initial *f*—*fibrek* ‘fabricated’, *flirtet* ‘flirted’, *filéax* ‘pilfered’, *fisfes* ‘missed’, etc.). When the derivation is automatic, as in the case of derived nominalizations (gerunds), analogical leveling with the verb is almost as likely to occur as within a particular verb paradigm (inflected Past and Future forms below are in 3rd person masculine singular):

	Past	Future	Imperative	Gerund
	<i>kiven</i>	<i>yexaven</i>	<i>kaven</i>	<i>kivun</i> ‘aim’
coll.	<i>xiven</i>	<i>yexaven</i>	<i>xaven</i>	<i>xivun</i>
or	<i>kiven</i>	<i>yekaven</i>	<i>kaven</i>	<i>kivun</i>
	<i>bikeš</i>	<i>yevakeš</i>	<i>bakeš</i>	<i>bikuš</i> ‘ask’
coll.	<i>vikeš</i>	<i>yevakeš</i>	<i>vakeš</i>	<i>vikuš</i>
or	<i>bikeš</i>	* <i>yebakeš</i>	<i>bakeš</i>	<i>bikuš</i>
	<i>pitéax</i>	<i>yefatéax</i>	<i>patéax</i>	<i>pitúax</i> ‘develop’
coll.	<i>fitéax</i>	<i>yefatéax</i>	<i>fatéax</i>	<i>fitúax</i>
or	<i>pitéax</i>	* <i>yepatéax</i>	<i>patéax</i>	<i>pitúax</i>

Exception: in the case of quadriliterals, *b*-*v* pairs are subject to opacity (and consequently to analogy) with the stop only, owing to the tendency to keep reduplicated syllables identical:

	Past	Future	Imperative	Gerund
	<i>bilbel</i>	<i>yevalbel</i>	<i>balbel</i>	<i>bilbul</i> ‘confuse’
coll.	<i>bilbel</i>	<i>yebalbel</i>	<i>balbel</i>	<i>bilbul</i>
	* <i>vilvel</i>	* <i>yevalvel</i>	* <i>valvel</i>	* <i>vilvul</i>

Across *binyanim* (verb patterns), analogy is likely if the relationship is reasonably automatic, as in active-passive pairs:

Pa'al Past	Nif'al Past	Substandard Var.
<i>katav</i> 'write'	<i>nixtav</i> 'be written'	<i>niktav</i>
<i>kavaš</i> 'conquer'	<i>nixbaš</i> 'be conquered'	<i>nikbaš/nikvaš</i>

but not when the derivation is more specialized and less automatic, as in a causality relationship:

Pa'al Past	Hif'il Past	Potential Variant
<i>katav</i> 'write'	<i>hixtiv</i> 'dictate'	* <i>hiktiv</i>
<i>kašal</i> 'fail (intrans.)'	<i>hixšil</i> 'fail (trans.)'	* <i>hikšil</i>

In non-automatic derivation, analogical leveling is highly unlikely:

Pa'al Past	Derived Noun	Potential Analogy
<i>katav</i> 'wrote'	<i>mixtav</i> 'letter'	* <i>miktav</i>
<i>patax</i> 'opened'	<i>maftéax</i> 'key'	* <i>maptéax</i>
<i>balat</i> 'projected'	<i>tavlit</i> 'relief'	* <i>tablit</i>

When the alternation affects the second radical of the root, analogical leveling with the stop is not that common: in Pa'al it only tends to occur in the imperative, probably because of the new imperative being derived from the future-used-imperatively (see § 17.3.2):

Formal Imperative	Future	Colloquial Imper.
<i>šxav</i> 'lie down!'	<i>tiškav</i>	<i>škav</i>
<i>švor</i> 'break!'	<i>tišbor</i>	<i>šbor</i>
<i>šfox</i> 'spill!'	<i>tišpox</i>	<i>špox</i>

with possible "spill-over" to the past stem in child language: *šabar* for *šavar* 'he broke', *šapax* for *šafax* 'he spilled' (but never **šakav* 'he lay down') and in the future of Nif'al in analogy with the past form:

Nif'al Past	Future	Colloquial
<i>niškax</i> 'be forgotten'	<i>yišaxax</i>	<i>yišakax</i>
<i>nišbar</i> 'be broken'	<i>yišaver</i>	<i>yišaber</i>
<i>nišpax</i> 'be spilled'	<i>yišafex</i>	<i>yišapex</i>

More commonly, however, analogical leveling in this position is with the fricative found in the base past-tense form of Pa'al:

Past	Future	Colloquial
<i>raxav</i> 'ride'	<i>yirkav</i>	<i>yirxav</i>
<i>šavat</i> 'strike'	<i>yišbot</i>	<i>yišvot</i>
<i>tafar</i> 'sew'	<i>yitpor</i>	<i>yitfor</i>

In the case of *p-f*, the colloquial variant with *f* almost totally replaces the one with the stop in some cases, and variants with *v* are more frequent than ones with *x*. Possible explanation: in this environment, *f* may only be derived

from *p*, and occurrences in this position of *v* from historical *w* are rare, whereas *x* from historical *ḥ* is fairly frequent, i.e. inapplication of analogy prevents (potential) additional ambiguity. Clearly, however, the avoidance-of-ambiguity explanation can only account for **some** cases. As in stem-initial position, if the derivational relationship is not automatic, there is (normally) no analogy:

Pa'al Past	Derived Noun	Potential Analogy
<i>zaxar</i> 'remember'	<i>mazkir</i> 'secretary'	* <i>mazzir</i>
<i>šavar</i> 'break'	<i>mašber</i> 'crisis'	* <i>mašver</i>
<i>šafax</i> 'pour'	<i>mašpex</i> 'watering can'	* <i>mašfex</i>

On the whole, then, it seems that analogy is affected by the frequency of "surface violations" to start with, owing to historical segment mergers, borrowings, etc.; that analogies with *f* and *v* are commoner because less potential ambiguity will result; and that the less inflectional and less automatic the derivation, the lower the likelihood of paradigmatic leveling. Spirantization is no longer a phonetically motivated process, and its application is severely restricted and certainly is not automatic. With the analogy factor, though, it has become fairly variable. Speakers are aware of the stop–fricative relationship.

17.6. Processes that use phonological information, but have always been morphological

17.6.1. Formation of segolate plurals

Segolate plural formation, as in

Singular	Plural	
<i>mélex</i>	<i>mlaxim</i>	'king'
<i>kéves</i>	<i>kvasim</i>	'sheep'
<i>kélev</i>	<i>klavim</i>	'dog'
<i>bóker</i>	<i>bkarim</i>	'morning'
<i>séfer</i>	<i>sfarim</i>	'book'
<i>kótel</i>	<i>ktalim</i>	'wall'

is often described as a phonological process, from the historical segolate base, as in

/malk+im/ 'king' > *malakim* (by *a*-insertion) > *mlakim* (by antepenultimate *a*-deletion) > *mlaxim*

As noted in § 17.5.1, the phonetic motivation for ante-pretonic *a*-deletion in MH is weak, and this ad hoc *a*-insertion cannot be motivated at all. Segolate

plural formation, though defined in phonological terms, is a truly morphological rule, paralleling broken plural formation in Arabic. Speakers simply know that the plural of masculine segolates is CCaC+*im*. This is a large, prominent class, and its deviant plural pattern is reinforced by numerous members, i.e. is not likely to disappear or to be leveled.

The situation is different in segolate **feminine** nouns. Alongside commonly heard normative plurals like

Sing.	Plural	
<i>giv(°)a</i>	<i>gva(°)ot</i>	'hill'
<i>dim(°)a</i>	<i>dma(°)ot</i>	'tear'
<i>yalda</i>	<i>yeladot</i>	'girl'
<i>ricpa</i>	<i>recafot</i>	'floor'
<i>simla</i>	<i>smalot</i>	'dresss'
<i>šimša</i>	<i>šmašot</i>	'pane'

one also finds non-normative forms based on the singular:

Sing.	Normative Pl.	Colloquial	
<i>darga</i>	<i>dragot</i>	<i>dargot</i>	'rank'
<i>kalba</i>	<i>klavot</i>	<i>kalbot</i>	'bitch'
<i>mišxa</i>	<i>mešaxot</i>	<i>mišxot</i>	'ointment'
<i>malka</i>	<i>mlaxot</i>	<i>malkot</i>	'queen'

It appears that speakers are not aware of the "segolate origin" of these feminine nouns, which look no different from any other feminine noun with a stressed feminine suffix. They simply know that **some** nouns with +a^f ending take the plural form CCaC+*ot*. It is also doubtful that speakers are aware of the non-segolate status of similar forms in which the initial #CV+ sequence is a prefix (usually ^f*mi+*/^f*ma+* or ^f*ti+*/^f*ta+*), which accounts for their **not** being realized as CCaC+*ot*:

Sing.	Plural	
<i>micva</i>	<i>micvot</i>	'commandment'
<i>mar'a</i>	<i>mar'ot</i>	'mirror'
<i>taxana</i>	<i>taxanot</i>	'station'
<i>tikva</i>	<i>tikvot</i>	'hope'
<i>tikra</i>	<i>tikrot</i>	'ceiling'
<i>misra</i>	<i>misrot</i>	'position'

They simply add +*ot*^f to the base, just as they do in /kalba+*ot*/ 'bitches' > *kalbot*.

The formation of segolate plurals, then, has always been morphological, and has nothing to do with phonetic naturalness. Speakers are clearly aware

of the relationships between the singular and plural patterns. If any variability is involved, it is restricted to some feminine plural segolates, which tend to be regularized in the colloquial register.

17.7. Conclusion

The phonetic–phonological–morphological continuum does not always work, but on the whole it provides a reasonably coherent picture of MH phonology, showing it to contain the whole gamut of processes and rules found in any other living language—even though it was “revived” as a living spoken medium after almost two millennia. They range from purely phonetic, automatic, variable processes, of which speakers are unaware; through phonological ones that are partially motivated phonetically and somewhat restricted morphologically; through rules that lost all phonetic motivation, have been morphologized, and have become quite opaque; to morphological rules that use phonological information but have always been morphological. With some notable exceptions, the continuum is also characterized by decreased variability and increased speaker awareness.

References

- Barkai, Malachi. 1978. “Phonological Opacity vs. Semantic Transparency: Two Cases from Israeli Hebrew.” *Lingua* 44: 363–78.
- Bat-El, Outi. 1989. *Phonology and Word Structure in Modern Hebrew*. Ph.D. dissertation, University of California, Los Angeles.
- Ben-Horin, Gad, & Shmuel Bolozky. 1972. “Hebrew *b*, *p*, *k*—Rule Opacity or Data Opacity?” *Hebrew Computational Linguistics* 5: 24–35.
- Blanc, Haim. 1964. “Israeli Hebrew Texts.” In *Studies in Egyptology and Linguistics in Honor of H. J. Polotsky*, ed. Haiim B. Rosén, pp. 132–52. Jerusalem: Israel Exploration Society.
- Bolozky, Shmuel. 1978. “Some Aspects of Modern Hebrew Phonology.” In *Modern Hebrew Structure*, ed. Ruth Aronson Berman, pp. 11–67. Tel Aviv: Universities Publishing Projects.
- . 1979. “On the New Imperative in Colloquial Hebrew.” *Hebrew Annual Review* 3: 17–24.
- . 1980. “Paradigm Coherence: Evidence from Modern Hebrew.” *Afroasiatic Linguistics* 7/4: 103–26.

- . 1982. "Remarks on Rhythmic Stress in Modern Hebrew." *Journal of Linguistics* 18: 275–89.
- . 1985. "The Domain of Casual Processes in Modern Hebrew." *Linguistic Analysis* 15: 19–27.
- Bolozky, Shmuel, & Adnan F. Haydar. 1986. "Colloquial Gender Neutralization in the Numeral System of Modern Hebrew and Lebanese Arabic." *Al-'arabiyya* 19: 19–28.
- Bolozky, Shmuel, & Ora R. Schwarzwald. 1990. "On Vowel Assimilation and Deletion in Modern Hebrew." *Hebrew Annual Review* 12: 23–48.
- Gil, David. 1986. "A Prosodic Typology of Language." *Folia Linguistica* 20: 165–231.
- Henkin, Roni. to appear. *The Imperative and Its Substitutes in Modern Colloquial Hebrew*. (Mediterranean Language Review supplementary series).
- Kadmon, Nirit. 1983. "On Main Stress, Secondary Stress, Stress Changes, and Casual Speech Deletion in Modern Hebrew." General paper, University of Massachusetts/Amherst, ms.
- Kiparsky, Paul. 1973. "'Elsewhere' in Phonology." In *A Festschrift for Morris Halle*, ed. Stephen R. Anderson and Paul Kiparsky, pp. 93–106. New York: Holt.
- McCarthy, John J. 1986. "OCP Effects: Gemination and Antigemination." *Linguistic Inquiry* 17: 207–63.
- Schwarzwald, Ora R. 1981. *Diqduq umeci'ut bapo'al ha'ivri*. Ramat Gan: Bar Ilan Press.
- Ulan, Russell. 1978. "The Nature of Future Tenses." In *Universals of Human Language*, ed. Joseph Greenberg, Charles A. Ferguson, & Edith A. Moravcsik, vol. 3, *Word Structure*, pp. 83–123. Stanford: Stanford University Press.
- Wexler, Paul. 1990. *The Schizoid Nature of Modern Hebrew: A Slavic Language in Search of a Semitic Past* (Mediterranean Language and Culture Monograph Series 4). Wiesbaden: Harrassowitz.

This page intentionally left blank.

Modern Aramaic Phonology

Robert D. Hoberman

State University of New York at Stony Brook

18.1. Introduction

Several entirely separate Aramaic languages are spoken today by a total of two or three hundred thousand people. The modern Aramaic languages fall into four distinct subfamilies, of which three are small both in the number of speakers and in the amount of dialectal diversity within the subfamily. These are: (1) the Ma'lūla group, spoken in three villages near Damascus; (2) the Tūrōyo group, including Tūrōyo proper, spoken in the Tūr 'Abdīn region of southeastern Turkey, and the language of the village of Mlaḥsô in east-central Turkey; (3) Mandaic, spoken in the city of Ahwaz, Iran, and perhaps also in Khorramshahr. These three subfamilies thus amount to four thoroughly distinct languages. (4) The fourth subfamily is modern Northeastern Aramaic. Spoken by the Christian and Jewish minorities of Kurdistan and Iranian Azerbaijan, that is, in northern Iraq and adjacent parts of Iran, Syria, and Turkey, it has far more speakers than all the other branches put together—possibly over two hundred thousand. It is impossible to say precisely how many distinct languages are in this subfamily, because of the vagueness of the distinction between language and dialect. Even so, it is realistic to say that modern Northeastern Aramaic comprises about six distinct languages, which are different in their grammatical structures and mutually comprehensible only with serious difficulty, if at all. Within modern Northeastern Aramaic the linguistic diversity, in terms of phonology, morphology, syntax, and vocabulary, is of the same order of magnitude as in all of colloquial Arabic. Together with the four languages of the other three subfamilies, the total number of modern Aramaic languages is thus about ten. In terms of time-depth of separation, the Ma'lūla group, representing western Aramaic, separated from all the others well over two thousand years ago. Modern Northeastern Aramaic and Tūrōyo bear significant similarities, and have been classed together by Jastrow (1990b) as two branches of a single subfamily, Central Neo-Aramaic.

This article will describe the phonologies of two very different Northeastern Aramaic languages, one in some detail and one mentioning just the highlights. For phonological information about the other modern Aramaic subfamilies, the reader can consult the following sources: on Ma'lūla: Arnold 1990; on Tūrōyo and Mlaḥsô: Jastrow 1985; on Mandaic: Macuch 1989 and the chapter by Malone in this volume.

18.2. Jewish dialects of northwestern Iraq

Within modern Northeastern Aramaic, the most conservative phonology is found in the Jewish dialects of northwestern Iraq, native to the towns of Zakho, Dohuk, Amadiya, Nerwa, Gzira, and others.¹ All the speakers of these dialects, aside from isolated individuals, are now in Israel. While these dialects differ among themselves in many details, they are all mutually comprehensible, their grammatical structures are the same in all fundamental respects, and their phonologies are likewise similar.² I take as representative of this group the dialects of the Jews of Zakho and Amadiya, abbreviated Z and Am. Previously published phonological studies of this group of dialects are Nakano 1969 and Hoberman 1989: 149–56.³ The following description complements the one in Hoberman 1989, presenting some of the same ideas and information in different ways, but some topics are not duplicated.

Most speakers of these dialects also spoke fluent Kurdish, the language of the surrounding Muslim majority, and their Aramaic has many loanwords and structural influences from Kurdish (and via Kurdish also from Turkish and Arabic). They also knew a certain number of Hebrew terms even before migrating to Israel, but only a few individuals knew Arabic.

The description that follows is synchronic. However, for the interest of those who are familiar with historical Aramaic, some basic diachronic information is also given. For more detailed information on the history of modern Aramaic phonology see Sabar 1976: xxxiii–xli, Nöldeke 1868, and Maclean 1895.

1. In the Aramaic dialect being described the names of these towns are *za:xo*, *dohuk* or *dho:k*, *ʕamidiya*, *nerwa* or *ne:rwa*, and *gzi:ra*. Amadiya in Modern Standard Arabic is *al-ʕama:di:ya*.

2. These dialects differ significantly from (but are mutually intelligible with) the Christian dialects of the same region, which are therefore not covered by the following description. Christian dialects of the region are documented in Krotkoff 1982 and Sara 1974.

3. In addition to numerous published Zakho texts, I have obtained a great deal of the Zakho data and some important observations from two sources for which I am especially grateful: unpublished notes of the late H. J. Polotsky, used with his permission, and extensive conversations with Professor Yona Sabar of UCLA, a native speaker of Zakho Aramaic.

Modern Aramaic sounds and words are written here in italics in a broad phonetic transcription equivalent to the American descriptivist phonemic level, which corresponds biuniquely with the phonetics. (A more convenient, simplified way of representing the vowels is suggested in § 18.2.4.) The hyphen is used to assist the reader by showing a morpheme boundary, but it has no phonetic implication. The symbol = indicates the linkage of a word and a proclitic or enclitic (see § 18.2.5.1). Occasionally phonetic or morpho-phonemic representations are used (in square brackets [] or vertical lines | | respectively), though even then the transcription is mainly phonemic in the above sense and phonetic detail or morphophonemic abstraction is shown only for the particular details being illustrated. Many words are identical in the Zakho and Amadiya dialects and where this is so no dialect labels are used. Thus a label Z or Am. implies that the word is different or lacking (or sometimes merely unattested) in the other dialect. There is one extension to this policy: since nearly all words with *θ* in Amadiya have *s* in Zakho, a word written with *θ* and appearing without a dialect label exists in Amadiya as given, with *θ*, and in Zakho in identical form but with *s*; for example, *be:θa* implies Am. *be:θa* and Z *be:sa*. Stress in words of more than one syllable is most often on the penultimate syllable, and will not be indicated in that case but only when it is other than penult.

18.2.1. Consonants

18.2.1.1. Inventory, phonetics, and phonological processes

Table 18-1 shows the consonant phonemes of the Amadiya dialect, excluding the emphatics, which are discussed below. Zakho has the same inventory

Table 18-1. Amadiya Aramaic Consonants

Bilabial		Labio-dental		Dental		(Alveo)-palatal		Velar		Uvular	Pharyn-geal	Glottal
p	b			t	d	č	ǰ	k	g	q	ʕ	ʔ
		f	v	θ				x	ɣ		ħ	
				s	z	š	ž					
	m				n							
					l							
					r							
			w				y					h

except that it lacks θ , while the dialect of Dohuk has both θ and δ .

18.2.1.1.1. PHONETIC SPECIFICATIONS

All the symbols have their conventional values, and only a few details need to be mentioned specifically. The voiced pharyngeal ʕ is a stop, probably a glottal stop with simultaneous pharyngeal constriction, unlike some varieties of Arabic in which it is a frictionless continuant or a semivowel. The voiceless stops are slightly aspirated before a vowel. Before a velar stop n is velar and before a uvular it is uvular. The tap r and trill rr are often retroflex (but without the lip-rounding of some varieties of English). w is most often labiodental, less often bilabial. The fricative v occurs only in words borrowed from Kurdish; it is rare in Amadiya, since many words which contain v in Zakho have w in Amadiya, as Z *zvîrre*, Am. *zwîrre* 'he turned around'.

18.2.1.1.2. EMPHATIC COARTICULATION

The "emphatic" coarticulation is identical, both phonetically and phonologically, to the same phenomenon which is familiar in Arabic. The articulatory and auditory nature of emphasis is complex: its articulation is a complex of pharyngealization, velarization, and sometimes also laryngeal elevation, labialization, and increased tension. Studies of Arabic, as well as of a different variety of modern Aramaic (Odisho 1988: 114–19), have consistently shown the central component to be retraction of the tongue root producing constriction of the pharynx. Neighboring consonants and vowels assimilate in emphasis to an emphatic consonant. Hence there is normally more than one emphatic segment in a word, e.g. Am. *pɑːliː* 'that he go out', but emphasis does not always pervade a word; for Am. *m̐leːle* 'he filled', informants reject the pronunciation **m̐leːle*. Emphasis also occurs in the absence of any emphatic consonant in the demonstratives Am. *ʔawɑːha* (masculine) and *ʔayɑːha* (feminine) 'that yonder'.

The presence or absence of emphasis is phonemic, as the following minimal and near-minimal pairs illustrate.

<i>p</i>	<i>gupaːle</i> 'crutches'	Am. <i>gupp̐algaːya</i> 'middle (adj.)'
		<i>gimp̐aːleʔ</i> 'he divides'
<i>b</i>	<i>baːʃ</i> 'good'	Z <i>baːʒir</i> 'city'
	Am. <i>baːʒir</i> 'city'	<i>baːθe</i> 'he will come'
		<i>baːte</i> 'houses'
<i>t</i>	Z <i>tʔaːya</i> 'to seek'	Z <i>tʔaːra</i> 'to wake up'
	<i>ʃtiːxa</i> 'spread'	Z <i>ʃtiːqa</i> 'quiet'
	Z <i>ʃtiːha</i> 'lying down'	
	Z <i>tʔreːle</i> 'he drove (an animal)'	<i>tre</i> 'two'

	<i>qitla</i> 'killing'	<i>Z qitil</i> 'murder'
<i>q</i>	<i>ʔo:da, ʔo:de</i> 'room, rooms'	<i>ʔo:da, ʔo:de</i> 'slave, slaves'
<i>ʂ</i>	<i>xa:ša</i> 'back'	<i>Z xa:sa</i> (Am. <i>xa:θa</i>) 'sister, new'
	<i>bišla</i> 'onion'	<i>pišra</i> 'meat'
<i>z</i>	<i>Z qa:ze</i> 'that he arrange'	<i>Z qa:ze</i> 'ducks'
	<i>Z qa:zax</i> 'that we arrange'	<i>Z qa:za</i> 'duck'
	<i>qaza:ne</i> 'pot' <i>Z</i> sg., Am. pl.	
	<i>Z qza:ya</i> 'to spend time'	<i>Z qza:la</i> 'neck'
<i>m</i>	<i>Z ma:ya</i> , Am. <i>ma:e</i> 'water'	<i>ma:yiθ</i> 'that he die'
<i>ç</i>	<i>Z çme:la</i> 'it was extinguished'	<i>Z jmidla</i> 'she froze'
	<i>Z čo:ʔa</i> 'smooth'	<i>čo:l</i> 'desert, wilderness'
	<i>Z çeʔle</i> 'he became smooth'	<i>Z çehya</i> 'tired (fem.sg.)'
	<i>Z çya:ʔa</i> 'be smooth'	
	<i>Z maço:ʔe</i> 'to smooth'	
<i>m</i> and <i>l</i>	<i>mle:le</i> 'he filled'	<i>mle:le</i> 'it sufficed'
	<i>gma:le</i> 'he fills'	Am. <i>gma:le</i> 'it is enough'
		<i>Z kma:le</i> 'it is enough'
	<i>Z mʂamo:re</i> 'bother, boss around'	<i>Z maʔmo:re</i> 'build'
	<i>Z la:ma</i> 'cheek'	
<i>r</i>	<i>Z kre:le</i> 'he rented'	<i>Z kre:le</i> 'he was/became short'
	<i>Z kiryā</i> 'rented (adj., masc.sg.)'	<i>Z kiryā</i> 'short (masc. sg.)'
	<i>kire</i> 'rent, fee, payment'	Am. <i>giran</i> 'expensive'
	<i>to:ra</i> 'Torah'	<i>to:ra</i> 'bull'
	Am. <i>gma:rix</i> 'he smells'	<i>gma:reʔ</i> 'it hurts'
	<i>Z qo:ri</i> 'kettle'	<i>Z qo:ri</i> 'my grave; they bury'
	Am. <i>ʔirxa</i> 'guest' (<i>Z ʔarxa</i>)	<i>ʔirxe</i> '(water-)mill'
	Am. <i>ʔurxa</i> 'road' (<i>Z ʔurxa</i>)	<i>ʔurza</i> 'male'
	Am. <i>ʔurwa</i> 'big'	<i>ʔirba</i> 'sheep, goats'
<i>r</i> and <i>p</i>	<i>Z rpe:le</i> '(an animal) was sicced (released to attack)'	<i>Z rpe:le</i> 'was loose'
	<i>Z marpo:ye</i> 'to sic (an animal)'	<i>Z marpo:ye</i> 'to make loose'
<i>p</i> and <i>l</i>	<i>palla</i> 'burning coal'	<i>palge</i> 'half, middle'
	<i>pli:ma</i> 'crooked, twisted'	<i>pli:xa</i> 'worked'
	<i>Z pallunka</i> 'crippled'	
	<i>Z lappa</i> 'paw'	
<i>p</i> and <i>z</i>	<i>po:za</i> <i>Z</i> 'face, countenance',	<i>po:xa</i> 'wind'
	Am. 'chin'	
	<i>Z zappa</i> 'large turd'	

<i>p</i> and <i>ç</i> Z <i>ça:pa</i> ‘slap’	Z <i>čippiksa</i> ‘drop’
Am. <i>čappa</i> , pl. -e ‘handful’	
Am. <i>čappe</i> (Z <i>čappe</i>) ‘left (side)’	
<i>r</i> and <i>z</i> Z <i>gzirre</i> (or <i>gzirre</i>) ‘he decreed’	Z <i>gzirre</i> ‘he circumcised’
Z <i>zaŋra</i> ‘huge’	Z <i>zo:ra</i> Am. <i>zŋo:ra</i> ‘small’
<i>č</i> and <i>t</i> Z <i>čaŋta</i> ‘suitcase’	Z <i>čanta</i> ‘cloth shoulder bag’

Consonants occur geminated intervocalically: *libba* ‘heart’, *xitte* ‘wheat’, *sippa:θa* ‘lips’ (sg. *sipθa*), *ŋatta* ‘now’, *ŋizza* ‘she-goat’, *mxaθθo:θe* ‘renew’, *gʷθθa*, pl. *gʷθθa:θa* ‘ball of yarn’, Z *miŋŋo:ja* ‘washed-out sliver of soap’, Z *mažža* ‘brain’, *ŋaxxa* ‘here’, *šinne* ‘years’, *šimma*, pl. *šimma:he* ‘name’, *gilla:le* ‘grasses, herbs’, *hayya* ‘quickly, early’, *hawwa* ‘good’, Am. *šawwarči* ‘photographer’; *šaŋŋa* ‘hour’, Z *rahhhat* ‘calm, relaxed, at ease, healthy’, Am. *šihha:ta* ‘matchbox’. The glottals *ʔ* and *h* are not attested as geminates. At the beginning of a word a geminate consonant appears when a prefix is identical (either underlyingly or by assimilation) to the first consonant of the stem: *p-pa:yiš* lb+pa:yišl ‘he will become or remain’ *b-ba:xe* ‘he will cry’, *m-ma:yiθ* lb+ma:yiθl ‘he will die’, *q-qa:re* lk+qa:rel ‘he studies’. Initial gemination is not always audible unless it is preceded by a vowel-final word.

Adjacent consonants assimilate regressively in voicing and emphasis, Am. lb+dabt+il *bdapti* ‘they will heal’, lgxik+lél *kxikle*, lb+qa:re:l *pqa:re* ‘he will read’, lbixza:yal, *biyza:ya* ‘seeing’ (cf. *xa:ze* ‘that he see’), *šlo:θa* ‘prayer’, Z *mista* lmizztal ‘hair’ (pl. *mizze*). In assimilation *ʔ* counts as voiced: *gʔa:riq* ‘he runs’. In addition the future-tense prefix *b* becomes *m* before a following nasal: lb+na:pil *mna:pil* ‘he will fall’, lb+ma:yiθl *mma:yiθ* ‘he will die’. The general present tense prefix *k-/g-* assimilates to an adjacent consonant in the usual manner illustrated above, but it also assimilates across an intervening vowel in verbs lacking an initial consonant, being voiceless in *ke:xil* ‘he eats’, *ke:θe* ‘he comes’, but voiced in *ge:mir* ‘he says’, Am. *ge:wid* Z *ge:wiz* ‘he does’, *ge:zil* ‘he goes’. (The future tense prefix does not alternate in this way, e.g. *ba:xil*.) The lexicalized nature of voicing in the *k-/g-* prefix is evident in Z *kma:le* ‘it suffices’ (in contrast to the usual case, where the prefix is voiced before nasals: *gma:le* ‘he fills’, *gma:yiθ* ‘he dies’, and Am. *gma:le* ‘it suffices’).

Like other consonants, *ʔ* can appear in any position within a syllable or word: *ʔa:la* ‘to spin (thread)’, *qaŋʔa* ‘that she cut’, *tarʔa* ‘door’, *paʔla* ‘laborer’, *qteʔle* ‘he cut’, *daʔra* ‘that she return’, *ša:meʔ* ‘that he hear’, *qa:teʔ* ‘that he cut’. Initially before a vowel there is no opposition between *ʔ* and zero, and *ʔ* is usually pronounced; thus *ʔa:zil* ([ʔa:zil] or [a:zil]) is ambiguous between lʔa:zil ‘that he spin’ (from the root *ʔzl*, historical **ʕzl*), and la:zil ‘that

he go' (from the root *zl* or $\emptyset zl$, historical $*?zl$); the difference is apparent in other forms: *b?a:zil* 'he will spin', *ba:zil* 'he will go', *?zilla* 'she spun', *zilla* 'she went'. Nonetheless $?$ may elide optionally, especially in a syllable coda, as in *ša:me?*, *qte?le*, for which *ša:me*, *qte:le* are also heard.

18.2.1.2. *Historical excursus*

More than two thousand years ago Aramaic underwent a sound change known as spirantization, in which *p b t d k g* changed into the corresponding fricatives when postvocalic and not geminate. These fricatives were at first allophones, but the loss of many short vowels and of some instances of gemination caused them to be phonemic and spirantization to be a morphophonemic rule. In modern Aramaic the rule is dead, all the words with a given root having either the stop or the fricative, for instance, from historical $*lbš$: *Iwa:ša* 'wearing, putting on (clothing)', *la:wiš* 'that he wear', *malwo:še* 'dressing'. There are only a few relic alternations, which are listed below. Hand in hand with the loss of alternation, subsequent sound changes affected three of the fricative variants, so that they are less similar to the corresponding stops than they were at first. Spirantized $*b$, which was at first undoubtedly a bilabial fricative, has merged with historical $*w$ as *w*, as in $*kātebā$ which became *kaθwa* 'that she write', or $*dābeq$ which is now *da:wīq* 'that he hold'; it frequently formed diphthongs, most of which are reduced to long vowels, as in *do:qa* *ldawq-al* 'that she hold', *ytu:la* *lyti:w+1+al* 'she sat down', *Z mut(t)o:ta* Am. *muttawta* 'placed, seated (fem.sg.)'. Spirantized $*g$ has merged with the reflexes of $*ʕ$ and $*ʔ$ as $ʔ$, such as *pe?la* 'radish'. Spirantized $*p$ has merged back with the stop as *p*, as in *na:piq* 'that he go out', *la:yip* 'that he learn'. However, it remained as *w* (more precisely, it is reflected in coalesced diphthongs *o:* or *u:*) in three words: *mno:š-* 'by oneself' from $*b+napš-$; *!lo:xe* 'lentils', cf. Syriac $*!lāphē$; *ru:ša* 'shoulder', cf. Syriac *rapšā* 'shovel'.⁴ The fricative variant of $*k$ remains as *x*: *xille* 'he ate'.

The interdentalals θ δ exist as such in the dialect of the Jews of Dohuk, one of the same group of dialects with Zakho and Amadiya, as well as in many Christian dialects. In Zakho these have become sibilants *s z*. In Amadiya $*\theta$ remained, while $*\delta$ became a stop, *d*. Speakers from Nerwa took the

4. On the etymology of *ru:ša*, a loanword from Akkadian, including the semantic connection of 'shovel' and 'shoulder', see Krotkoff 1985: 126–27. The change of *t* to *w* also applied in a couple of loanwords, Am. *tāwsir* 'translation of a religious text' (from Arabic *tafsīr*), Am. *tawtiš* 'inspection' (from Arabic *taftīš*).

Amadiya pattern one step further, shifting the θ to s .⁵ The correspondences, to which there are only very rare exceptions, are thus as follows:

	'house'	'hand'
Dohuk	<i>be:θa</i>	<i>ʔi:ða</i>
Zakho	<i>be:sa</i>	<i>ʔi:za</i>
Amadiya	<i>be:θa</i>	<i>ʔi:da</i>
Nerwa	<i>be:sa</i>	<i>ʔi:da</i>

Older Aramaic $\hbar$ and ς shifted to x and ʔ respectively, merging with the reflexes of spirantized $*k$ and $*\text{ʔ}$. Pharyngeal $\hbar$ and ς are pronounced, however, in numerous words of Arabic origin, some of them otherwise completely Aramaicised, such as *maḥko:ye* 'speaking'. It would be incorrect to say that $\hbar$ and ς were first lost and then restored as borrowed phonemes; in reality there was never a stage of the language that lacked $\hbar$ and ς . This is true because they survive in a few native Aramaic words, especially in the environment of emphatics and q : *raḥu:qa* 'far' and the verb *rḥa:qa* 'be far', *rumḥa* 'spear', *ʕuṭma* 'hip', *ʕapša* 'gall nut', *mʕaʕo:te* 'farting', *ʕamu:qa* 'deep'. They are also pronounced in numerous borrowings from Hebrew, and such borrowing must have been taking place continuously since ancient times. In any case, $\hbar$ and ς are plentiful and fully integrated in the modern language.

18.2.1.3. Relics of spirantization

Frozen relics of the formerly productive (originally allophonic) process of spirantization remain in a few lexicalized alternations.

$b \sim w$: *za:win* 'that he buy', *mza:bin* 'that he sell'; *kalba* 'dog', pl. *kalwe*; *ʔirba* 'sheep', pl. *ʔirwe*; *xša:wa* 'to think', *xižbo:na* 'thought'; and perhaps b 'in', *xa-w-ʔiṣra* 'ten times as much'.

$g \sim \text{ʔ}$ or ς : *palge* 'half', *mpalo:ʔe* 'to split'; *ga:w-* 'in', *l-ʕo:ya* 'inside' (cf. *l-warya* 'outside' from $*l\text{-bar}$).

$t \sim \text{Am. } \theta, Z s$: *be:θa*, 'house', pl. *ba:te*. There are two partially rule-governed patterns of alternation. (1) The two variant forms of the feminine ending *-ta* $\sim$ *-θa* (Z *-sa*) are selected partly on a lexical basis,⁶ but in deverbal nouns and adjectives they are governed by a simple rule: the suffix is *-ta* after conso-

5. An identical asymmetry between the reflexes of $*\theta$ and $*\delta$ is found in the dialect of the Jews of Azerbaijan (Garbell 1965), in which $*\delta$ became d but $*\theta$ became l , remaining a continuant (thus *ida* 'hand', *bela* 'house'), and also in the Ashkenazic pronunciation of Hebrew, with s for $*\theta$ but d for δ ... *yod* 'hand', *bayis* 'house'. Jewish dialects of southeastern Kurdistan have l for both. The Christian Urmi dialect and Iraqi Koine have stops: *be:tʰa*, *i:da*.

6. Sara 1993 analyzes the complex patterning of these two variants in a closely related dialect.

nant-final stems (e.g. Am. *ptixta* ‘open’) and after the vowels *o:* and *u:* derived from coalescence of a vowel with a final *w* from historical spirantized *b*: Am. *kθu:ta* (historical **kθībtā*) ‘written’, but it is *-θa* after vowel-final stems derived from final-*y* roots: *gli:θa* ‘revealed, exposed’. (2) A second fairly regular process changes stem-final *θ* (and its reflex *s* in Zakho) to *t* before a suffix beginning with *l*. This happens in preterite forms of the verb *mya:θa* ‘die’, such as *mitle* *lmiθ+lel* ‘he died’; in Zakho the intervening suffix *-wa:-* ‘past’ prevents this, so *mīswa:le* ‘he had died’, but in Amadiya the *t* is characteristic of the preterite stem even with *-wa:-*, thus Am. *mītwa:lu* ‘they had died’. Similarly Am. *ʔi:θ Z ʔi:s* ‘there is’, *ʔitle*, *Z ʔīswa:le* ‘he had’, for which Amadiya has both *ʔīθwa:le* and *ʔītwa:le*.

18.2.2. Syllable types

The basic syllable types, which may appear in any part of a word, are: CV, CV:, and CVC.⁷ At the beginning of a word an extra consonantal position may be added to any of these, allowing for an unlimited variety of initial consonant clusters.⁸ These are so common and varied primarily because the most common verb template is of the form CCV:C. At the end of a word an extra consonantal position may also be added, although this is relatively rare, yielding two additional syllable types: CV:C and CVCC. Of these two, CVCC is the rarer, occurring in about 1% of the vocabulary, all in words borrowed from Arabic and Kurdish, including Am. *dawramand* ‘wealthy man’,

7. In Amadiya, where intervocalic *ʔ* and *y* before front vowels tend to elide, there are a number of sequences of vowels and thus of medial syllables beginning with a vowel: *ki:e* ‘he knows’, *ki:in* ‘I (masc.) know’, *ki:a* ‘she knows’ (for *ki:ʔe* etc.), Am. *ma:e* ‘water’ (for *ma:ye*), and in general all sequences of *a:ye* or *a:yī*, as in many verbs like *pa:yīš* ‘become, remain’, may be pronounced [a:e], [a:i], also *ʔamo:(y)i* ‘my paternal uncle’ for *ʔamo:y-i*, cf. *ʔamo:yux* ‘your paternal uncle’.

The rare tense vowels in closed medial and closed unstressed final syllables are phonetically rather short, as in *Z brindar* ‘wounded’ (cf. *bri:n* ‘wound’), Am. *zilga* ‘a match (for lighting)’, *nerwa* (name of a town), *gmenxa* ‘she looks’ (cf. *gme:nix* ‘he looks’), Am. *jemke* ‘twins’, *bemke* ‘pine nuts’, Am. *ʔitwa* (in free variation with *ʔītwa* ‘there was’, *bista:na* ‘garden’, Am. *fista:na* (but *Z fista:na*) ‘woman’s dress’, Am. *qurduθkī* ‘in Kurdish’, *siruθkī* ‘in Christian Aramaic’ (cf. *qurduθ* ‘the Kurds, Kurdish language’, *siruθ* ‘Syriac Christendom’), *taxmin* ‘thought’, *hakim* ‘physician’. It is difficult to determine whether there is a contrast of vowel length between words like *ne(:)rwa* or *be(:)mke* and words like Am. *herge* ‘clothing’, Am. *meθya* ‘that she bring’, since there are few words of either class and the nature of the following consonants, whether sonorants or stops, seems to play a role.

8. Detailed listings of consonant clusters in other dialects are given in Sara 1974: 41–47 and Odisio 1988: 64–78.

Am. *dirist* ‘straight’, Z *xurʔ* ‘violent’, *raʃt* ‘right’, *bass* ‘only’, Z *miʃʃ* ‘fog’, Z *sirr* ‘deep cold’. Final syllables of the form CV:C exist only when stressed, and occur productively in the imperative singular and (in Amadiya only) the bare preterite of first conjugation verbs and in certain inflectional suffixes. Examples: *bri:n* ‘wound’, *ba:ʃ* ‘good’, *čo:l* ‘wilderness, desert’, *ye:r* ‘other’, *hi:l* ‘until’, *he:ʃ* ‘yet, still’, *ʔi:θ* ‘there is’, Am. *li:t* Z *le:s* ‘there is not’, *ha:l* ‘situation’, *ʃwa:n* ‘young’, *ʃwi:n* ‘side, party’; for verbs see the following section.

18.2.3. Canonical shape and minimal length of words

Native Aramaic nouns and adjectives are at least of the structure CVCCV or CV:CV, and there may be additional preceding material: Am. *šimša* ‘sun’, Am. *garma* ‘bone’, Am. *ʔuʔma* ‘thigh’, Am. *gupta* ‘cheese’, Am. *ke:pa* ‘stone’, Am. *go:ra* ‘man’, Am. *ga:re* ‘roof’, Am. *ʔirxe* ‘mill’, *ʔiro:ta* ‘Friday’, *ʔili:θa* ‘fat tail of sheep’, Am. *raki:xa* ‘soft’, Am. *ʃaʔmo:ta* ‘pomegranate’, Am. *qinya:na* ‘large domestic animal’, Am. *guppalba:ya* ‘middle (adj.)’. Loanwords may be shorter, and may end in a consonant: *giran* ‘expensive’, *čo:l* ‘wilderness’, Am. *gaha* ‘knuckle’, *kire* ‘fee, payment’.

The syllabic shape of verbs is strictly governed by the canonical forms of the system of discontinuous (nonconcatenative) morphology in the usual Semitic mold, in which the shortest regular verb forms, first conjugation imperatives, have the shape (C)CV:(C): *šqo:l* ‘take’, *xo:l* ‘eat’, *ʔzi:* ‘see’. A few irregular imperatives are shorter: *hal* ‘give’, *mar* ‘say’, *θa* ‘come’.

Adverbs, prepositions, exclamations, and other closed-class items are frequently very short: *biʃ* ‘more’, *ču* or *čʰ* ‘no’, *ʔe(:)* ‘this’, *xu* (exclamation of surprise), *dʔ* (exclamation of offering, giving: ‘here!’), *ba* (exclamation: ‘but, so’), *b* ‘with, at, by’, *d* ‘that, of’, *w* or *u* ‘and’.

18.2.4. Vowels

18.2.4.1. General properties

There are seven distinct vowel qualities, *i i e a o u*, and a feature of vowel length, V:. These are organized into a system of five long and three short vowel phonemes: *i: e: a: o: u: i a u*; there are in addition a few mostly rule-governed occurrences of short *i*, *e*, *u*, and *o*. *i* and *u* are always short, while the others—the five peripheral vowel qualities—occur phonetically both long and short. However, the only vowel quality for which there is a secure phonemic opposition of long and short quantity is *a*. For the other four peripheral vowels, *i e o u*, the long and short variants are in complementary distribution; they are short in closed non-final syllables and in unstressed

final syllables whether open or closed, and long otherwise; details will be specified below.⁹ In the position of maximal inventory of vowel oppositions, open non-final syllables, there are thus eight phonemic vowels, five long and three short: *i*: *e*: *a*: *o*: *u*: *i* *a* *u*. In the position with the smallest inventory, non-final syllables closed by an obstruent (other than glottal), there are three vowels, all short: *i* *a* *u* (plus short *i* and *u* in a handful of items each). In final open unstressed syllables, there are five vowels: *i* *e* *a* *o* *u*. For details of these inventories, see Hoberman 1989: 153–54.

The transcription used here could be simplified without loss of information in the following way. Instead of *i* and *u* write *i* and *u*, and write [i] and [u] always with the sign of length, even when they are short but tense: *i*: and *u*:. This causes no confusion because [i] and [i:], [u] and [u:] are in complementary distribution, as mentioned above. Thus the words *zilla* ‘she went’, Am. *zilga* ‘match’, *bista:na* ‘orchard’, *hakim* ‘physician’, *ʔurza* ‘male’, *siruθ* ‘Syriac Christendom’, *si*: ‘go (imperative)’, *di* (an interjection associated with offering) would be written as *zilla*, *zi:lga*, *bi:sta:na*, *haki:m*, *ʔurza*, *siru:θ*, *si*:, *di*. The sign of length can be omitted in unstressed final vowels, where all vowels are short except in certain intonation contours (discussed below) and *i* and *u* do not occur: *kxa:ze* ‘they see’, *kθu:li* ‘I wrote’. Although for the purpose of this article I chose to use a more phonetically revealing system, the simpler system just described is phonemically adequate and would be suitable for such purposes as the publication of texts, lexicons, and grammatical analysis.¹⁰

18.2.4.2. Vowel quality

The vowels *i* and *u* are more central and distinctly lower than *i* and *u*, *i* varying from [ɪ] to [ə] and *u* from [ʊ] to [ə].¹¹ Thus *i* and *u* differ mainly in terms of rounding, rather than in the front–back dimension, while the contrast between *i* and *i*(:), and similarly between *u* and *u*(:), is more in respect to quality than it is in quantity. The phonemic opposition between *i* and *u* is

9. There is a marginal phonemic opposition of quantity in the cases of *i* and *e*. In *xam-šī=ba:te* ‘fifty houses’ the stressed *i* is short but tense, as opposed to long *i*: in *si*: ‘go (imperative)’ and to lax *i* in *dī* (an interjection associated with offering). In *mēθi:li* ‘bring me (imperative)’ the *e* is usually short, as opposed to long *e*: in *me:θi* ‘bring (imperative)’. Short *e* and *o* are discussed in more detail below.

10. This system was used in Avinery 1988, following unpublished work by H. J. Polotsky. A different sort of simplification, taking advantage of the low functional load of vowel quantity to reduce the number of symbols, was used in Hoberman 1989, following Nakano 1969.

11. Because of this central quality some authors have written *u* as *ü*.

demonstrated by such pairs as *ʔirba* ‘sheep, goats’ vs. *ʔurza* ‘male’, Am. *ʔurwa* ‘big’, Am. *witwa* ‘you (masc.sg.) were’ vs. Am. *wutwa* ‘you (pl.) were’, *ʃargume* ‘turnips’ vs. *ʔimme* ‘thin, flat breads’, *gilla* ‘herb, grass’ vs. *ʃulle* ‘clothing’, Am. *ʔirxa* ‘guest’ vs. *ʔurxa* ‘road, way’. However, in some environments the opposition is neutralized, adjacent labial and emphatic consonants tending to produce a sound that is more like *u*, as in *ʃruxlu* for *ʃri:x+lul* ‘they shouted’, and there is sometimes free variation, as between *ʔidyo* and *ʔudyo* ‘today’, *pumma* and *pimma* ‘mouth’, *ʕi=mindi* and *ʕu=mindi* ‘nothing’.

Some speakers pronounce *o*: as distinctly fronted, sometimes almost to the point of IPA [ø], as in [pø:ʃi:ya] ‘turban’, Z [pø:xa] ‘wind’. This fronting is prevented by a neighboring emphatic, [po:za] Am. ‘chin’, Z ‘face and upper body’, but not by *q*: [ʃø:qi:] ‘that they leave’, Am. [pø:qa] ‘nostril’. *u*: is not similarly fronted: [gø:ra] ‘man’, [gu:re] ‘men’.

Long *a*: and short *a* are low central vowels in most environments, low back in the neighborhood of emphatics. Some Zakho speakers front short *a* [æ] before an *r* in the same syllable, provided it is not preceded by an emphatic: [kærma] ‘orchard, vineyard’, [qærsa] ‘cold’, but [ʔarpa] ‘leaf’. In Amadiya final *a* is strikingly rounded and raised to [ɔ] regardless of the preceding consonant. In Amadiya too *a*: is rounded and raised to [ɔ:], almost [o], next to an emphatic labial consonant, as in [bɔ:ʃ] ‘good’ (but Zakho [bɑ:ʃ]), Am. [bɔ:ʒir] ‘city’ (Z [ba:ʒir]), Am. [gmɔ:le] ‘he fills’, Am. [mɔ:li] ‘my property’, but no rounding in Am. [gma:le] ‘it is enough, stop!’, [ba:be] ‘his father’, [ʔa:le] ‘to him’, Am. [gma:rix] ‘he smells’.

Short *e* and *o* are only marginally phonemic, as nearly all of their occurrences are predictable by three simple and superficial rules:

1. All word-final unstressed vowels are short, though tense and peripheral, including *e* and *o* as in Am. *de:re* ‘monastery’, Am. *ka:lo* ‘bride’, Am. *ʃe:qo* ‘trouble’, *kxa:ze* ‘he sees’.
2. Before *ʔ* or *h* we find *e* and *o* instead of *i* and *u* respectively. This rule is quite regular in verb conjugation and in deverbal nouns and adjectives, so that corresponding to *ptixle* ‘he opened’ we have *ʃmeʔle* ‘he heard’.¹² Similarly *e* and *o* appear before *h* or *ʔ* in other nouns and adjectives: *behna* ‘breath, moment’, *dehwa* ‘gold’, *sehra:ne* ‘communal festive picnic’, *beʔta*

12. It apparently does not apply in preterite and perfect forms of second-conjugation verbs with a first root consonant *ʔ*: Z *mʔimra* ‘built’, Am. *mʔwiʔte* ‘he caused to enter’, *mʔu:ʔa:le* ‘he caused her to enter’.

‘egg’, *beʔe* ‘eggs’, Am. *šmoʔta* ‘fame’, *zdoʔθa* ‘fear’, Z *čoʔta* ‘smooth (fem.sg.)’.¹³

3. Short *e* and *o* appear in a few morphemes, especially in Zakho, as a free variant of the more frequent *i* and *u*. All of these alternate morphologically with *e:* and *o:* respectively. These items are the following:
 - the suffix *-et* ~ *-it* ‘second person masculine singular’ in *ʔa:het* ~ *ʔa:hit* ‘you (masc.sg.)’, *wet* ~ *wit* ‘you are’, which has morphologically conditioned or free alternates *-e:tin* (verbal agreement marker) and plural *-e:tu:n*;
 - the second person masculine singular possessive suffix Z *-ox*, sometimes *-o:x* (Am. *-ux*), plural *-o:xu:n*;
 - the extended diminutive suffix *-un-ka* ~ *-on-ka* ~ *-in-ka* ~ *-o:n-ka*, e.g. Z *ʔwanqonkat* ‘lads’, the first part of which appears alone in historical diminutives such as *bro:na* ‘son’;
 - the enclitic form of *xe:ta* ‘other’, which may be pronounced *xe:t*, *xet*, or *xit*.

Aside from these rule-governed instances of *e*, there are a few items showing that it is a phoneme marginally distinct from other vowels. They are *miθya* ‘brought (participle, masc.sg.)’ versus Am. *meθya* (Z *masya*) ‘that she bring’ (cf. *me:θe* ‘that he bring’); Am. *begwi:ne* (also Am. *bəgwi:ne*) but not **bigwi:ne* ‘eyebrows’ versus *bigwa:ra* ‘marrying’;¹⁴ Am. *ʔeyya* ‘this (masc. and fem.)’ vs. *hayya* ‘quickly’. It is clear from all this that the phonemic status of *e* is more secure than that of *o*.

18.2.4.2.1. DIPHTHONGS

The older Aramaic diphthongs *ay* and *aw* generally coalesced to *e:* and *o:* respectively in these dialects. However, they exist in a small number of words, many of them borrowed, with considerable dialectal variation: Am. *kawdīnta* ‘mule’ (Z *ko:zinta*), Am. *ʔawra:za* (Z *ʔavra:za*) ‘hill’, Z *gawda* ‘trunk (of body)’, Am. *muttawta* (Z *mit(t)o:ta*) ‘having been seated, placed’, Z *xawʔisra*

13. In Amadiya some words show free variation: Am. *dehna* and *dihna* ‘oil rendered from meat’, *šeʔde* and *šiʔde* ‘almonds’ (in Zakho only *dehna*, *šeʔze*), but not all do: Am. *šuʔla* ‘work; thing, matter’ does not have the alternate **šoʔla* (cf. Z *šu:la*). Before pharyngeals and liquids the vowel sometimes sounds more like *e* than like *i*, as in *mo:ʃel* ‘Mosul’, *mfo:ʃel* ‘cut’, Am. *ba:zer* ‘city’, Am. *mo:deʃlu* ‘they lost’, Am. *pi:ramer* ‘old man’, but they are in free variation with *i*. Both Am. *herge* and *hirge* ‘clothing’ are heard (but in Zakho only *hirge* ‘ragged, poor clothing’).

14. *begwi:ne* was etymologically a compound, **be:(θ)* + *gwi:ne*, cf. classical Syriac *ghīnē* ‘eyebrows’.

‘ten times as much’, Am. *tawsir* (Z *tafsir*) ‘translation of a religious text’, Am. *tawtfš* ‘search, examination’, Z *ʔaw* ‘that (masc.)’, *ʔay* ‘that (fem.)’, *ʔayda* ‘benefit, use’, *mayle* (or *ma=yile*) ‘what=is’. Several forms of the verb Am. *msaho:we* contain the diphthong *ew*: *gimsa:hew* ‘he gets frightened’, *mso:hewli* ‘I got frightened’ (in Zakho the verb is not used and the related noun Z *sahve* ‘fear’ and adjective *sahva:na* ‘awesome’ have *v*).

18.2.4.3. Vowel quantity

Long *i*: *e*: *a*: *o*: *u*: are longest in open syllables and when stressed, shorter in closed syllables and when unstressed. Unstressed at the end of a word, such as the *i* of *kšaqli* ‘they take’, vowels are normally short, though tense. This is true whether the word is in the middle of a phrase or at the end of a sentence, but at the end of a phrase or clause which does not end a sentence, i.e. in a situation which would call for a comma in English orthography, such vowels are long.

Vowel quantity is phonemic, and it is not too hard to find minimal or near-minimal pairs to prove it: *bale* ‘but’ vs. *ba:le* ‘his attention’; Z *bala* ‘trouble’ vs. *ba:la* ‘attention’; *ma:ne* ‘that he/they count’, *mani* ‘who’; *gibe* ‘he wants’ vs. *ge:be* ‘his direction’; Am. *šargume* ‘turnips’ vs. Am. *gu:ma* ‘hole, pit’ vs. *ko:ma* ‘black’; *me:θi* ‘bring (imperative singular)’ vs. *méθi:li* ‘bring me’. However there are strong statistical correlations between vowel length and syllable type: open syllables tend to have long vowels, and closed syllables tend to have short vowels. Thus the functional load of vowel quantity in modern Aramaic is rather low. If we define UNMARKED QUANTITY as long in open syllables and short in closed syllables, then only about 8% of all syllables in a running text have vowels of marked quantity.¹⁵ To see the significance of this we can compare it with another Semitic language which likewise has an opposition of vowel quantity and which moreover belongs to the same *Sprachbund* as the Aramaic dialects being described: the Arabic spoken by the Jews of Aqra and Arbil (Jastrow 1990a). In this Arabic dialect, marked syllables make up about 40% of the total: five times the proportion in the Aramaic of Zakho and Amadiya.¹⁶ We can go further by

15. For this purpose final unstressed vowels (as in *baxta* ‘woman’ or *šaqli* ‘that they take’) are counted as long. Although their phonetic length is variable, as described above, they are tense, not lax. Moreover they are long when any suffix is added: *šaqli:le* ‘that they take him’.

16. The distinctiveness of modern Aramaic vis-à-vis languages of the world lies in the unusually high frequency of long vowels in Aramaic. Thus the text frequency of all long vowels in a Zakho or Amadiya Aramaic text is about 64% of all vowels (or 29% if we exclude final unstressed tense vowels such as that in *šaqli*). Compare this with 45% (or 31%)

looking at specific types of syllables. Of marked syllables, the most prevalent type is an open syllable with a short vowel. These amount to about 5% of all syllables. Of marked syllables in which the vowel is long, the majority (2% of all syllables) are stressed monosyllabic words, such as *čə:l* 'wilder-ness', *dmo:x* 'go to sleep (imperative)'.

The preponderance of what I have called unmarked syllables—closed syllables with short vowels and open syllables with long vowels—is supported by numerous and productive alternations of the type *kša:qil* 'he takes', *kšaql-a* 'she takes'. The proper analysis of such alternations is not certain, however, because there are many other types of alternations between long and short vowels and there are a fair number of deviations from this pattern of alternation. Long and short vowels alternate in the following kinds of situation. Note that in most of these alternations *i* is the short counterpart of both *i:* and *e:*, and *u* is the short counterpart of both *u:* and *o:*.

1. Open versus closed syllable with syncope or epenthesis of *i*, V:CiC (or V:CiC-C) ~ VCC-V. This pattern is extremely common, being totally productive in verb inflection. Examples: Am. *kpa:tix* 'he opens', *kpa:tixlu* 'he opens them', *kpatxa* 'she opens', *kpatxa:lu* 'she opens them', *kpatxin* 'I open', *kpatxe:tu:n* 'you (pl.) bring'; Am. *mo:rimle* 'he lifted, raised', *mum:ile* 'he raised them', *ke:xil* 'he eats', *kixli* 'they eat', *kixle:tu:n* 'you (pl.) eat', *gme:θe* 'he brings', *gme:θax* 'we bring', Am. *gmeθya* Z *gmasya* 'she brings'.

COUNTERINSTANCES: Several borrowed nouns and adjectives of the same shape do not lose the *i* or shorten the vowel when a suffix is added: *ša:qil* 'mind', *ša:qile* 'his mind', *ša:qilo:xun* 'your (pl.) mind' (all with the usual penultimate stress). Several native Aramaic prepositions exhibit similar syncope or epenthesis of *i* but the vowel is always short: *baθir* 'after', *baθri* 'after me', and the construct states of many nouns of the shape CVCCV, where the third consonant is *n*, *l*, or *r*; are similar: *pisir* 'meat of' (from *pisra* 'meat'); *šadīr* 'chest of' (from *šadra* 'chest').¹⁷

2. Open versus closed syllables, with vowel-initial or consonant-initial suffixes respectively, V:C-V, VC-C. This pattern too is common and productive in several noun and adjective forms, including participles and verbal

in the Arabic of Aqra and Arbil, or with 33% (or 14%) in Cairo Arabic. Greenberg 1966: 18–20) examined seven languages with an opposition of vowel quantity and found that long vowels were from 8 to 25 percent of all vowels in a text.

17. It would be easy to distinguish these three types (*ša:qil/šaqla*, *ša:qil/ša:qile*, *baθir/baθri*), and non-alternating words like *raš*, *rašte* 'left' too, by means of distinct abstract underlying shapes and diacritic rule features, but the correct approach is not obvious and this is not the place to work out all the details.

nouns. Examples: *pti:xa:le* 'he opened her/it', *ptixle* 'he opened'; *pti:xa* 'taken (masc.sg.)', *pti:xe* 'taken (pl.)', *ptixta* 'taken (fem.sg.)'; *smo:qa*, *smuqta* 'red (masc., fem.)', *ruqta* '(an act of) spitting', *ro:qe* 'saliva', pl. *ruqya:θa*; Am. *ši:da:na*, *ši:danta*, Z *šiza:na*, *šizanta* 'crazy (masc., fem.)', *hako:ma* 'king', *hakumta* 'queen', *rahu:qa*, *rahuqta* 'far (masc., fem.)'.

3. Second-conjugation verbs with single middle consonants (historical *aʿel*) and those with two middle consonants (historical *paʿel*) differ in the vowels of the jussive, preterite, and imperative stems. The one-consonant shape is C(C)V:CVC, as in *ma:lip* 'teach', *mša:dir* 'send', *na:bił* 'take'; the two-consonant shape is (C(C)VCCVC, as in *manxip* 'shame', *mašmiš* 'fondle', *mparčik* 'pluck (a fowl)', *mxallil* 'wash'. These differences correlate exactly with the open or closed nature of the first syllable:

'teach' 'shame'

gma:lip *gmanxip* 'he teaches, shames'

mo:lipte *mūnxipte* 'he taught, shamed'

4. Long vowel when stressed, short when unstressed (both in open syllables). This pattern is productive in certain categories of second conjugation verbs, such as Am. *mšo:dir* 'sent (preterite)', *mšudarta* 'sent (stative participle, fem.sg.)'. It also applies to *a:* in a number of nouns, both native and borrowed, such as Am. *ʔa:xa* (Z *ʔaxo:na*) 'brother', pl. *ʔaxawa:θa*; *ja:sus* 'spy', pl. *jasu:se* 'spies'; *za:xo* 'Zakho', *za(:)xo:na:ya* 'of Zakho (adj.)', Z *su:ris* (Am. *siruθ*) 'Syriac Christendom', *sura:ye* '(Aramaic-speaking) Christians'. In nominals (nouns, adjectives, and infinitives) only *a:* alternates in this manner, other vowels remaining constant whether stressed or unstressed: *ʔi:la:na* 'tree', *ʔi:za:la* 'going', *po:ši:ya* 'turban', Am. *be:da:θa* 'long decorative sleeve cuffs'.

COUNTERINSTANCES: There are words which retain a long vowel even when not stressed, such as *ba:ba* 'father', pl. *ba:bawa:θa*. Here the first *a:* is phonetically shorter than the second, stressed one, but yet not as short as the fully short first vowel in *ʔaxawa:θa* 'brothers'. This retention of vowel length in unstressed position is productive in verb forms containing one or more of the unstressable suffixes (discussed in § 18.2.5), thus Am. *kpa:tix* 'he opens', *kpa:tixle* 'he opens it', *kpa:tixwa:le* 'he used to open it'; Am. *mo:qid* '(they) burnt him, he was burnt', *mo:qidle* 'he burnt'. These cases can be analyzed as involving cyclic stress assignment and vowel lengthening (Hoberman 1989: 116–18).

5. The short vowel of a final closed syllable corresponds to a long vowel when the stem has a vowel-initial suffix, VC# ~ V:C-V. This happens in both stems and suffixes. Examples: Am. *ba:žir* 'city', pl. *ba:že:re*; *ja:sus* 'spy', pl.

ḡasu:se; Z *ḡe:wan* ‘animal’, pl. *ḡe:wa:ne* (Am. *ḡaywan*, *ḡaywa:ne*); *kūḡḡmaṛ* Am. ‘high-ranking Christian cleric, priest’, Z ‘habitually angry person’, pl. *kūḡḡma:ṛe*; Am. *qa:zan* ‘cooking pot’, pl. *qa:za:ne*; *gma:lāš* ‘he dresses (transitive)’, *gmālu:ša* ‘she dresses’ (on the shortening of the *a* see paragraph 7 below); *kpalxīt* ‘you (masc.sg.) work’, *kpalxe:tun* ‘you (pl.) work’, Am. *plixlūx* Z *plixlox* ‘you (masc.sg.) worked’, *plíxlo:xun* ‘you (pl.) worked’. A similar case is *ḡi:la:na* ‘tree’, *ḡi:lān* ‘tree of’ (construct state). In the subject-agreement suffixes of verbs, first and second person singular, there are pairs with no difference of meaning, of the following types, illustrated by forms for ‘I/you work’: *kpalxin* / *kpalxe:na* (1.masc.), *kpalxan* / *kpalxa:na* (1.fem.), *kpalxīt* / *kpalxe:tin* (2.masc.), *kpalxat* / *kpalxa:tin* (2.fem.). There are nouns and adjectives, especially but not only of the pattern |CaCi:Cl, such as *ḡakim* ‘physician’, pl. *ḡaki:me*, *sarbor* ‘anecdote, incident’, pl. *sarbo:re*, in which the final stem vowel, when suffixless, is unstressed and short, but not centralized; similarly a suffix forming abstract nouns has the two forms *-uθ* and *-u:θa* (the distribution is lexical), as in *xli:muθ* ‘thickness’, Am. *sīruθ* ‘Syriac Christendom’, (also *sīruθkī* ‘in Christian Aramaic’), versus *hawu:θa* ‘good(ness), benefit’.

6. Monosyllabic words of the form C(C)V:C shorten the vowel when a consonant-initial suffix is added. This occurs in first-conjugation imperatives and (in Amadiya but not Zakho) preterites, and in forms of the existential predicator *ḡi:θ*. Examples: *pto:x* ‘open (imperative)’, *ptūxle* ‘open (imperative) it’; Am. *pti:x* ‘(they) opened it, it was opened (preterite)’, *ptixle* ‘he opened’; *ḡi:θ* ‘there is/are’, *ḡitle* ‘he has’, *ḡiθwa* ‘there was/were’, Z *le:s* Am. *li:θ* ‘there is/are not’, Z *laswa* Am. *liθwa* ‘there was/were not’, Z *latle* Am. *litle* ‘he doesn’t have’.¹⁸ All of these monosyllables have alternative forms, with no discernable difference of meaning, with an added *-in*: *pto:xin*, *pti:xin*, *ḡi:θin*, *le:sin* / *li:tin*, which has the effect (we might even say purpose) of eliminating the occurrence of a marked closed syllable with a long vowel, creating an unmarked open syllable. In Zakho this *-in* is obligatory with unsuffixed monosyllabic preterites.¹⁹

7. In a stressed antepenultimate syllable the vowel is often short. Examples: *ptūxu:le* ‘open (pl.) it’ (cf. *pto:xun* ‘open (pl.)’), *ptūxle* ‘open (sg.) it’;

18. In Amadiya all the forms of the existential predicator with *θ* can also be pronounced with *ɪ*.

19. This *-in* is similar in its apparent meaninglessness to the *-in* that attaches optionally to the first and second person subject markers mentioned just above, where it does not have the same effect of opening a long-voweled syllable. An apparent difference in function between forms with and without *-in* is discussed in Sabar 1976: 40 n. 34.

kpatxítu:le ‘you (pl.) open it’ (cf. *kpatxe:tun* ‘you (pl.) open’); *gmáhi:bin* ‘I like’ (cf. *gma:hib* ‘he likes’); *méθi:le* ‘bring him (imperative)’ (cf. *me:θi* ‘bring (imperative)’).

COUNTERINSTANCES: *mšá:dirre* ‘send (imperative) him’ and all such imperatives of second-conjugation verbs with a single middle consonant.

8. As mentioned above, final unstressed vowels are normally short. If suffixes are added, in most cases these vowels are lengthened (or it may be said that they are underlying long and shorten in final position): *kšaqli* ‘they take’, *kšaqli:lu* ‘they take them’.

9. In a clitic pair, if the first element ends in a stressed vowel it is usually short: *go:rá=go:ra* ‘(man=man) = each man’; *la=θe:le* ‘(not=he.came) = he did not come’; *la=ba:š* ‘no=good’; *xamšá=ba:te* ‘five=houses’; *čθ=na:ša* ‘no=person’. In Amadiya the first consonant of the second element of the clitic pair is sometimes geminated after the short stressed vowel (thereby creating an unmarked syllable type): Am. *čθ=gga* ‘never’, Am. *čθ=xxa* ‘no=one’, Am. *gθ=ttu:ra* ‘in mountain’. This has been lexicalized in the adjective Am. *gup-palga:ya* ‘middle’, from *gu* ‘in’ + *palg-* ‘half’ + *-a:ya* (adjectivalizing suffix).

Several typical situations in which marked syllables are found have been mentioned up to this point. In addition to those already mentioned, short vowels in open stressed syllables occur in many loanwords, such as *šargume* ‘turnips’, include those with the feminine suffix *-ita* (stress on the penult as usual; from Arabic *-at* via Kurdish or Turkish), such as *xilmita* ‘work’, *qu:wi-ta* ‘strength’, and also in a few native words such as *gibe* ‘he wants’, *mare* ‘possessor of, one with’, *mani* ‘who’.

18.2.5. Stress and clitic units

Stress in words of more than one syllable falls most often on the penultimate syllable. Throughout this chapter stress is indicated only when it is other than penultimate. In this section, however, stress is indicated in all cases.

Stress is phonemic, as the following minimal pairs demonstrate: *bále* ‘but’ versus Z *balé(:)* ‘certainly’, *má:liple* ‘teach him (imperative singular)’ versus *ma:liple* ‘that he teach him’ (and similarly with all verbs of the second conjugation). However, stress shifts fairly readily for emphasis, either a vague semantic emphasis or to emphasize the length or shortness of a vowel for the benefit of the inquiring linguist (stressed long vowels are longer and short vowels more clearly short than unstressed).

In words other than verbs, stress is uniformly penultimate. This is true even in loanwords where the stress in the source language is in another position (except for the most unassimilated, conscious borrowings), such as

the following words from Arabic: Am. *ḥákim* ‘physician’, Am. *ḥáywan* ‘animal’, Am. *májbūr* ‘compelled’ (but with an enclitic copula *májbūr=i:win* ‘I am compelled’), Am. *samáwi* ‘sky blue’, *quwwáta* ‘strength’; and from Hebrew: *ba:ruxxábba* ‘welcome’. There are a small number of lexical exceptions, almost all of them adverbs: Am. *derḥál* ‘thereupon’, Am. *hammán* ‘slightly, hardly’, Am. *dúrtidyom* ‘on the next day’ (Z *dírtyom*), Am. *š-xáflati* ‘suddenly’, Am. *ṭamá(:)* ‘why’, Am. *siruθkí* ‘in Syriac (Christian Aramaic)’, Am. *qurduθkí* ‘in Kurdish’, Am. *bínafši* ‘purple’, *l-báxxati* ‘on my conscience, mercy’, Am. *tawtíš* (Z *táftiš*) ‘search, examination’, Am. *ʔárbu:šub* (Z *ʔar-bó:šib*) ‘Wednesday’.

In verbs stress is governed by morphological rules, though even here the penultimate pattern is still evident as the default. In general, a verb form has the following morphological makeup:²⁰

(TENSE.ASPECT)–STEM–(GNP)–(*wa:*)–(*I*-GNP)–(*I*-GNP)

The following rules specify the place of stress in verbs:

1. Imperative forms (which have no prefix) have initial stress:

Am. *šqál-u:-l-i* ‘take (plural) me’, Am. *má:líp-l-e* ‘teach (sg.) him’, Z *máʔu:r-u:-l-e* ‘bring (pl.) him in’, Z *máýze:-l-e:-l-i* ‘show him to me’.

2. Otherwise stress is as close to the penult as permitted by the following restrictions:

(a) In certain second-conjugation verbs a penultimate vowel *i:* or *u:* is a non-syllabic semivowel at a more abstract level of analysis; this syllable does not count in stress placement, and the words consequently have antepenultimate stress: Am. *g-málu:š-a* |g-malwš-al ‘she dresses’ (derived from *lwa:ša* ‘to wear’), *g-máhi:b-in* ‘I like’ |g-maḥyb-inl. When further suffixes are added the stress shifts: *g-malu:š-á:-l-e* ‘she dresses him’, *g-maḥi:b-in-n-u* (l-in-l-ul) ‘I like them’.

(b) Stress does not fall on suffixes of either of two classes: the past tense marker *-wa:-* and the set of GNP markers which are preceded by *-l-* (though stress can freely fall on suffixes of the other GNP set). If the word contains one or more such suffixes amounting to two or more syllables, stress will be earlier than

20. “GNP” stands for a suffix marking the gender, number, and person of the subject, direct object, or indirect object of the verb; there are two different sets, one of which is always preceded by the element *l*. The suffix *wa:* is a past tense marker. Various combinations of prefixes and suffixes are permitted in various tense/aspect categories, but their order is always the same.

penultimate: *šqíl-wa-l-ux* ‘you had taken’ *šqíl-l-o:xu:n* ‘you (pl.) took’. Occasionally in continuous speech (though not in citation) this prohibition is violated.

(c) The second syllable of the second person plural suffix *-e:tu:n ~ -ítu:-* is never stressed: *k-šaqł-é:tu:n-wa* ‘you (pl.) would take’, *k-šaqł-ítu:-l-e* ‘you (pl.) take him’.

(d) Otherwise stress is penultimate: Am. *k-pá:tíx* ‘he opens’, *k-patx-é:tun* ‘you (pl.) open’, *gi-mšádr-ax* ‘we send’, *gi-mšádr-é:tun* ‘you send’, *qam(m)šádr-áx-l-u* ‘we sent them’, *k-pa:tíx-wa* ‘he would open’, *k-patx-í:-wa* ‘they would open’.

18.2.5.1. Clitic groups and stress shift

In certain kinds of two-word collocations the normal stress on the second word is reduced or eliminated, the main stress being on the first word.²¹ This is indicated by the symbol = between the two words. Such cliticization occurs in phrases of the the following types:

1. The present positive copula is normally enclitic to the noun, adjective, etc. that precede it, as in *dawrámand=i:le* ‘he is a rich man’. In Amadiya words that end in *-a* (i.e. most nouns) have that vowel coalesce with the vowel of the copula, as in *kaθá:wa* ‘scribe’ with *i:le* ‘he is’, pronounced [kaθáwe:le] ‘he is a scribe’ (Z [kasá:wale]). In such cases there is no fully satisfactory place to write the = boundary, because to write *kaθa:we=le* incorrectly implies that the noun is the plural *kaθá:we* (cf. *kaθa:we=lu* ‘they are scribes’). I opt for *kaθa:we=le* nonetheless, because it correctly indicates the pronunciation, including the stress; *kaθá:w=e:le* would be more misleading, because it would suggest that the location of the stress is irregular. After vowels other than *a* the *i:* is absent: *ká:lo=la* ‘she is a bride’.
2. Verbs preceded by the negative marker *la*, such as *lá=kšaqłìn-wa:lu* ‘I would not take them’, or by the negative copula, *le:we=gwí:ra* [lé:wegwì:ra] ‘he is not married’, and generic nouns with the negative marker *čú ~ ču*, such as *čú=míndi* ‘nothing’ or with *kull-* or *kud* ‘all’: Am. *kúlle=míndi* ‘everything’.

21. Such a collocation is termed a “secondary stress unit” by Krotkoff (1982: 16–17) and a “Betonungseinheit” or “Akzentkompositum” by Jastrow (1985: 29–32, 1988: 20–21).

3. Participles or infinitives followed (in Amadiya but not Zakho) by the object marker =*ll*- (a reduced form of the preposition *ʔill*-) plus GNP suffixes: *bi-yzá:ya=llux* ‘seeing you’.
4. Numerals with following enumerated noun: *xamší=bà:te* ‘fifty=houses’. The stress in the numeral generally shifts to the final syllable.
5. Reduplications: *he:dí=hè:di* ‘slowly’, *go:ṛá=go:ṛa* ‘(man= man) = each man’, *ṭu:ré=ṭu:re* ‘full of mountains, along the mountains’, *xabrá=p=xabra* ‘word by word’, *kò:la:ná=b= ko:là:na* ‘street by street’. In these, too, stress shifts to the end of the first member.

18.3. The Christian dialect of Urmi and Koine

The Aramaic dialects described above are phonologically the most conservative of modern Northeastern Aramaic. To illustrate the degree of variation found in this subfamily I will take the dialect of the Christians of Urmi, Iran, and its offshoot, the Modern Aramaic Koine of the Assyrians in Iraq. This is probably the most widely spoken Aramaic language today. The phonology of these varieties has been described before (Marogulov 1935/1976, Polotsky 1961, Hetzron 1969, and especially Odisho 1988), so I will merely mention four important innovations.

18.3.1. *The distribution of emphasis*

In these varieties of Aramaic a word is either emphatic or plain as a whole. Thus emphasis affects all the sounds of the word *.muddírtela* ‘she has given back’ (it is indicated here by the dot before the word), which thus contrasts as a whole with *muddírtela* ‘she has dared’ (Hetzron 1969: 113). There are only very limited exceptions, such as *i.ša:rat* ‘sign, signal’, in which *i* is plain but *.ša:rat* is emphatic. The entire phenomenon, sometimes called “synharmonism,” has been very well described in both its synchronic and its diachronic aspects by Jušmanov (1938), and its articulatory and acoustic phonetics are treated in detail in Odisho 1988: 114–19.²²

18.3.2. *Aspiration*

The voiceless stops and affricates *p t č c* (also the rare *k*, but not *q*) are distinctively aspirated or unaspirated, with minimal pairs like *.pa:rɪ* ‘odd

22. The same phenomenon exists in the Jewish dialect of Urmi and the surrounding area. The phonology of emphasis in this dialect has been analyzed in detail in Hoberman 1988.

numbers' vs. *.p^ha:ɾɪ* 'lambs'. This is the case in both emphatic and plain words. There is thus a six-way opposition for stops and affricates, illustrated by *dya:la* 'giving birth', *tyara:ya* 'a member of the Tiari tribe', *t^hya:wa* 'sitting', *.dya:ra* 'returning', *.tya:p^ha* 'bending', *.t^hla:ma* 'punishing'.

18.3.3. *Fronting of uvulars and velars*

In the Urmi dialect *k*, *g*, and *q* have shifted forward, the velars to palatal *c* *j* (remaining distinct from the affricates *č* *ǰ*) and the uvular to *k*. In the Iraqi Koine *q* remains uvular while the velars are shifted to the palatal position, with the result that velar *k* and *g* exist only in a handful of baby-talk words and recent loanwords (Odisho 1988: 46).

18.3.4. *Vowel quantity*

The functional load of vowel quantity has been much reduced. Vowel length is almost entirely determined by syllable type and stress, although there are a few syllable types in which both long vowels and short *ɪ* can appear.

References

- Arnold, Werner. 1990. *Das Neuwestaramäische*, vol. 5: *Grammatik* (Semitica Viva 4). Wiesbaden: Harrassowitz.
- Avinery, Iddo. 1988. *Ha-niv ha-Arami shel Yehudey Za'kho* (The Aramaic dialect of the Jews of Zakho). Jerusalem: Israel Academy of Sciences and Humanities.
- Garbell, Irene. 1965. *The Jewish Neo-Aramaic Dialect of Persian Azerbaijan* (Janua Linguarum Series Practica 3). The Hague: Mouton.
- Greenberg, Joseph H. 1966. *Language universals with special reference to feature hierarchies* (Janua Linguarum Series Minor 59). The Hague: Mouton.
- Hetzron, Robert. 1969. "The morphology of the verb in Modern Syriac (Christian colloquial of Urmi)." *Journal of the American Oriental Society* 89: 112–27.
- Hoberman, Robert D. 1988. "Emphasis harmony in a modern Aramaic dialect." *Language* 64: 1–26.
- . 1989. *The syntax and semantics of verb morphology in modern Aramaic: A Jewish dialect of Iraqi Kurdistan* (American Oriental Series 69). New Haven.: American Oriental Society.
- Jastrow, Otto. 1885. *Laut- und Formenlehre des neuaramäischen Dialekts von Midin im Tūr 'Abdīn*, 3rd ed. Wiesbaden: Harrassowitz.

- . 1988. *Der neuaramäische Dialekt von Hertevin (Provinz Siirt)*. Wiesbaden: Harrassowitz.
- . 1990a. *Der arabische Dialekt der Juden von 'Aqra und Arbīl*. Wiesbaden: Harrassowitz.
- . 1990b. "Personal and demonstrative pronouns in Central Neo-Aramaic." In *Studies in Neo-Aramaic*, ed. Wolfhart Heinrichs (Harvard Semitic Series), pp. 89–103. Atlanta: Scholars Press.
- Jušmanov, N. V. 1938. "Singarmonizm urmiskogo narečija." *Pamjati akademika N. Ja. Marra*, pp. 295–314. Moscow & Leningrad: Akademija Nauk SSSR, Institut Jazyka i Myšlenija.
- Krotkoff, Georg. 1982. *A Neo-Aramaic dialect of Kurdistan* (American Oriental Series 64). New Haven: American Oriental Society.
- . 1985. "Studies in Neo-Aramaic lexicology." In *Biblical and related studies presented to Samuel Iwry*, ed. Ann Kort and Scott Morschauser, pp. 123–34. Winona Lake, Ind.: Eisenbrauns.
- Maclean, Arthur John. 1895. *Grammar of the dialects of vernacular Syriac*. Cambridge: Cambridge University Press. Repr. Amsterdam: Philo, 1971.
- Macuch, Rudolf. 1989. *Neumandäische Chrestomathie mit grammatischer Skizze, kommentierter Übersetzung und Glossar*. Wiesbaden: Harrassowitz.
- Marogulov, Q. I. 1935/1976. *Grammaire néo-syriaque pour écoles d'adultes (dialecte d'Urmia)*, trans. Olga Kapeliuk (Comptes Rendus du Groupe Linguistique d'Études Chamito-Sémitiques, Suppl. 5). Paris: Geuthner.
- Nakano, Akio. 1969. "Preliminary reports on the Zaxo dialect of Neo-Aramaic: Phonology." *J. of Asian and African Studies* (Tokyo) 2: 126–42.
- Nöldeke, Theodor. 1868. *Grammatik der neusyrischen Sprache*. Leipzig: Weigel. Repr. Hildesheim: Olms, 1974.
- Odisho, Edward Y. 1988. *The sound system of modern Assyrian (Neo-Aramaic)* (Semitica Viva 2). Wiesbaden: Harrassowitz.
- Polotsky, H. J. 1961. "Studies in modern Syriac." *Journal of Semitic Studies* 6: 1–32.
- Sabar, Yona. 1976. *Pəšaṭ Wayəḥî Bəšallāḥ: A Neo-Aramaic midrash on Be-shallāḥ (Exodus)*. Wiesbaden: Harrassowitz.
- Sara, Solomon I. 1974. *A Description of Modern Chaldean* (Janua Linguarum Series Practica 213). The Hague: Mouton.
- . 1993. "Marked gender in Modern Chaldean [ta/θa] suffix." In *Semitica: Serta philologica Constantino Tsereteli dicata*, ed. R. Contini et al., pp. 299–308. Turin: Zamorani.

This page intentionally left blank.

La phonologie des langues sudarabiques modernes

Antoine Lonnet et Marie-Claude Simeone-Senelle

C.N.R.S. et Université de Paris 3 - Sorbonne Nouvelle

19.1. Présentation

Au sud de la Péninsule arabique, à l'ouest du Sultanat d'Oman, à l'est de la République du Yémen et dans les îles yéménites du golfe d'Aden, vivent quelque deux cent mille Arabes dont la langue maternelle n'est pas l'arabe mais une des langues sémitiques que la tradition scientifique européenne a nommé *sudarabique moderne* (désormais: SAM).¹

Cette appellation pourrait malencontreusement laisser entendre qu'il s'agit de dialectes arabes; or il n'en est rien. L'intercompréhension n'est pas possible entre le sudarabique et l'arabe même si l'un et l'autre appartiennent, avec le *sudarabique ancien* (épigraphique) et les langues sémitiques d'Éthiopie, à la branche méridionale du sémitique de l'ouest. Le groupe SAM est subdivisé en six langues: mehri, harsusi, bathari, hobyot, jibbali et soqotri (phonétiquement, [mehri], [ħarsūsi], [baṭṭari], [ħəwbyūt], [ǧibbāli] et [soqotri]).

19.1.1. Localisation (se reporter à la carte)

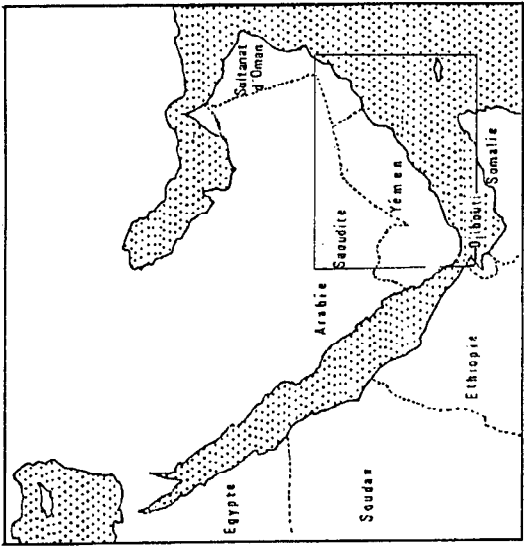
Les langues SAM sont parlées aujourd'hui dans deux pays.

a) Au Yémen:

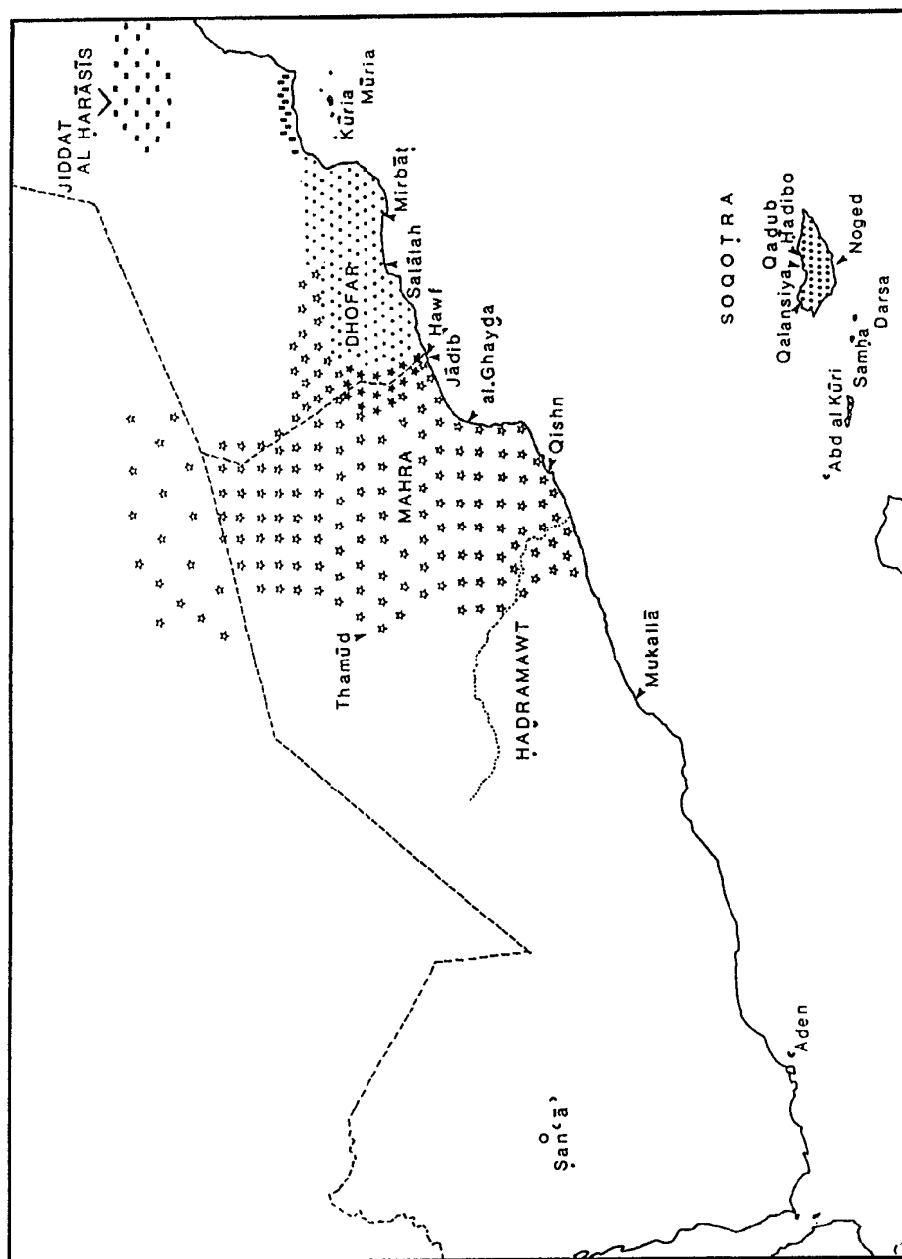
- le mehri est parlé dans le gouvernorat le plus oriental, celui du Mahra;
- le hobyot est parlé sur une aire réduite, à l'extrême est du Mahra, à la frontière du Sultanat d'Oman;
- le soqotri est parlé dans l'île de Soqatra et les petites îles voisines de 'Abd-al-Kūrī et Samḥa (l'îlot de Darsa n'est pas habité).

1. La recherche de terrain menée par la Mission Française d'Enquête Linguistique au Sud-Yémen (MFELSY), aujourd'hui Mission Française d'Enquête sur les Langues du Yémen, fondée en 1982 par les auteurs, est soutenue par le Ministère des Affaires Étrangères (Direction de la coopération Scientifique et Technique), l'Université d'Aden, le C.N.R.S. et l'Université de Paris 3 - Sorbonne Nouvelle. Cet article est adapté des chapitres correspondants de notre ouvrage à paraître *Les langues du pays de l'encens: le sudarabique moderne*.

					
Mehri	Harsusi	Bathari	Hobyot	Jibballi	Soqotri



Map 2. Les langues sudarabiques modernes



b) Dans le Sultanat d'Oman:

- le mehri est parlé à l'ouest, dans les montagnes du Dhofar;
- le jibbali est parlé dans la même région jusque dans les villes côtières et dans les îles Kuria Muria;
- le hobyot est parlé dans la montagne près de la frontière du Yémen;
- le harsusi est parlé dans la zone nommée Jiddat al-Ḥarāsīs;
- le bathari est parlé sur la côte, face aux îles Kuria Muria.

c) Il faut ajouter que nombre d'émigrés dans les pays du Golfe et en Afrique de l'est constituent des communautés qui ont conservé leur langue.

19.1.2. *Situation linguistique*

19.1.2.1. *Les langues sudarabiques et l'arabe*

La langue sudarabique est la langue première des communautés sudarabophones. Les locuteurs l'utilisent dans la vie domestique et dans toutes les relations sociales, et parfois même administratives, internes à la communauté.

Dans les régions limitrophes entre les langues sudarabiques, et dans les régions où plusieurs coexistent, la règle est le plurilinguisme: il n'est pas rare de rencontrer des personnes parlant trois langues sudarabiques en plus de l'arabe.

Dans le Mahra, au Yémen, plusieurs langues exercent une pression à des degrés divers: à l'ouest, le Hadramout est une aire de langue arabe et, à l'est, au Dhofar (Sultanat d'Oman), coexistent le jibbali, le mehri et l'arabe (sans compter le hobyot de la région frontalière).

L'arabe, langue officielle, est la langue seconde apprise à l'école et utilisée dans la majorité des activités de la vie de citoyen. Les progrès de la scolarisation et le développement des moyens de communication font que la quasi-totalité des sudarabophones de moins de quarante ans comprennent et parlent l'arabe. Seuls certains adultes âgés (surtout des femmes) des régions reculées connaissent encore très mal l'arabe. Il faut préciser que l'intercompréhension est exclue entre le soqotri et les langues SAM "continentales". Parmi celles-ci, le jibbali est difficilement compréhensible aux locuteurs des autres langues, c'est-à-dire du mehri et des trois langues qui en sont très proches. Dans de telles conditions, l'arabe, langue nationale, est amené à servir de langue de relation.

19.1.2.2. *Écriture*

Aucune de ces langues n'a de tradition d'écriture; quelques expériences montrent que l'alphabet arabe permet d'écrire le mehri moyennant un aménage-

ment très simple. Plus difficile est l'adaptation au jibbali; le soqotri noté en lettres arabes nécessite une *scriptio plena* très lourde: les (nombreuses) voyelles doivent être notées.

19.1.2.3. *Dialectes*

Les langues sudarabiques ont une riche dialectologie, dont un aperçu peut transparaître dans les exemples qui illustrent le présent article, volontairement et explicitement choisis dans la plus grande variété dialectale.

Pour se limiter au plan phonétique, nos enquêtes sur le terrain yéménite ont permis de dégager pour le mehri des critères de classement: coalescence dentales-interdentales, vocalisation de la fricative laryngale /ʕ/, timbres des voyelles et des diphtongues, glottalisation ou laryngalisation des emphatiques.

Dans les dialectes soqotri: coalescence des fricatives vélaires et pharyngales, durée des voyelles, voix murmurée, labialisation des voyelles postérieures, degré de palatalisation des vélaires, sonorisation des emphatiques, réalisation fricative de /l/.

En hobyot, il semble que malgré le très petit nombre de locuteurs, des différences dialectales assez importantes apparaissent.

Pour le jibbali, Johnstone (JL) distingue le dialecte de l'est et celui du centre, ne faisant qu'évoquer celui de l'ouest et celui des îles Kuria Muria.

19.1.3. *Découverte des langues sudarabiques modernes*

Au dixième siècle de notre ère al-Hamdani indique dans sa *Description de l'Arabie*:² "les Mahra ont un parler barbare, on dirait des étrangers". Le terme Mahra désigne évidemment pour lui toutes les populations de langue SAM.

Les savants arabes n'ignoraient donc pas l'existence des langues sudarabiques vivantes de leur époque mais la science linguistique arabe, pour des raisons historiques et culturelles, ne s'est pas intéressée à d'autres langues que l'arabe. On peut donc véritablement parler de découverte lorsqu'au dix-neuvième siècle la science occidentale constata que ces langues étaient toujours parlées.

Cette découverte s'est faite, souvent au gré du hasard ou de l'aventure, en plusieurs épisodes dont le plus fructueux reste l'Expédition Sudarabique de 1898.

2. Hamdani 1884: 134 (éd. Müller), ou Hamdani 1983: 248 (éd. al-Akwa').

Les premières informations linguistiques sur les langues SAM apparaissent au dix-neuvième siècle. J. R. Wellsted publie une liste de mots soqotri et mehri (1835 et 1840). F. Fresnel donne une brève description du jibbali, sous le nom de “*ehbkili*” (1838). Puis le corpus sudarabique s’enrichit des contributions de quelques voyageurs, navigateurs, explorateurs, missionnaires, savants, répertoriées par Leslau (1946), et, en 1898, l’Académie Impériale des Sciences de Vienne organise une expédition scientifique de grande envergure sous la direction conjointe de C. de Landberg et de D. H. Müller. Moins de trois ans après le retour de l’Expédition, l’ethnographe autrichien W. Hein est envoyé par l’Académie dans le Mahra pour compléter les études sudarabiques.

Ces voyages et ces collaborations ont permis de constituer une véritable bibliothèque sur les langues SAM. Les prestigieux volumes édités entre 1900 et 1909 par l’Académie dans la collection *Südarabische Expedition*, les comptes rendus de séances de l’Académie, les communications, les articles de revues scientifiques, sont les fondements des études sudarabiques: plus d’un millier de pages constituant aussi les premières archives de ces populations sudarabiques et une image de leurs sociétés il y a un siècle.

Les langues traitées étaient donc les mêmes que celles qui avaient été découvertes à partir de 1834: soqotri, mehri, jibbali (nommé *šhauri*).

M. Bittner, dès 1908, sut tirer de ces matériaux des études comparatives et descriptives dont une grande partie reste encore valide aujourd’hui, et, jusqu’à très récemment, c’est toujours l’Expédition sudarabique qui a alimenté les travaux en linguistique SAM, parmi lesquels les deux ouvrages fondamentaux: *Lexique soqotri* de W. Leslau (1938) et *Syntax der Mehri-Sprache* d’E. Wagner (1953).

Les autres langues sont découvertes au vingtième siècle: B. Thomas révèle en 1929 le harsusi et le bathari, deux dialectes mehri ayant connu une évolution particulière. Il en publie un corpus en 1937, avec du mehri et du jibbali (sous le nom de *shahari*).

L’existence du hobyot au Dhofar est évoquée par T. M. Johnstone en 1981; le premier article présentant le hobyot parlé au Yémen est publié par la MFELSY en 1985.

La fin du vingtième siècle est marquée par les importants travaux, en particulier les dictionnaires harsusi, jibbali et mehri (du Dhofar), de T. M. Johnstone, les travaux d’ethnolinguistique soqotri de V. V. Naumkin et V. Ja. Porxomovskij et dernièrement les travaux linguistiques d’A. Lonnet et

M.-Cl. Simeone-Senelle (MFELSY). D'autres chercheurs ont abordé le SAM: Ch. Matthews, A. Nakano et M. Morris.

19.1.4. Dénominations des langues

19.1.4.1. Le groupe sudarabique moderne

Ce groupe de langues aujourd'hui appelé le plus couramment *sudarabique moderne* ou *néo-sudarabique* (Neusüdarabisch, Modern South Arabian) a parfois reçu d'autres noms qu'il est utile de signaler: Contemporary South Arabian, (néo-)himyarite, mehritique, sudarabe, Modern South Arabic, et même arabe méridional (!).

19.1.4.2. Le jibbali

C'est le nom le plus récent et désormais le plus général pour cette langue; il succède à *šhauri* qui, choisi à tort par D. H. Müller en 1907, a servi pendant plus d'un demi-siècle à la désigner chez les orientalistes. Or les véritables noms de la langue étaient déjà connus à cette époque: *šhari*, *qarāwi*, *eḥkili*.

Il faut attendre 1969–70 et les articles de Ch. Matthews puis T. M. Johnstone pour avoir quelques éclaircissements sur ces noms.

Ġibbālī (*jabbāliyah* pour Matthews) et *šaḥri* sont deux traductions dans le dialecte arabe local de *šheri* 'montagnard, paysan', dérivé de *šher* (arabe local *šīhr*) 'montagne, campagne'. Le mot 'jibbali' a l'avantage d'être dépourvu de la connotation péjorative de *šheri* qui s'applique à un homme sans statut tribal, un serf. On retient donc, avec T. M. Johnstone, le terme qui ne signifie que 'langue de la montagne'.

La langue est aussi parlée par la tribu dominante dont le nom arabe est *qarā* (singulier *qarāwi*) et le nom jibbali *əḥklo* (singulier *əḥkli*).

En définitive les diverses appellations de la langue, sans préjuger de la diversité dialectale qui apparaît dans les diverses publications, sont les suivantes:

ġibbālī, *ġabbāliyah*, *ġəblēt*; *šheri*, *šheri*, *šīhri*, *šherēt*, *šaḥri*, *šīhri*, *shahari*; *əḥkli*, *eḥkili*, *ḥakili*; *qarāwi*, *gawī*, *qarauwi*, *grauwi*; sans compter une appellation géographique plus vague, *ḍofāri*.

Quant à *šhauri* (*shkhawri*, *šxawri*), c'est ainsi que l'informateur de D. H. Müller nomma un jour cette langue, expliquant que le mot signifie 'pauvre, faible', par opposition à 'tribal'. Ce mot n'est pas usité, mais on peut penser (Leslau 1938: 211) à la racine sudarabique <šxr>: mehri 'vieillard', jibbali 'vieillard, faible, pauvre', soqotri 'pauvre, faible, homme, individu'; le statut non tribal est dit 'faible', arabe *ḍaʿīf*.

19.1.4.3. *Les autres langues*

- Pour le *mehri*, [məhri], signalons que le nom mehri de cette langue est, en fonction de la morphologie du dialecte, *məhrīyət* à l'ouest du Mahra, *məhriyōt* à l'est, et *məhrāyyət* au Dhofar.
- Le *soqotri*, [soqotri], est désigné en soqotri par une périphrase: "la langue de Soqotra [səḵḷtri]".
- Le *harsusi*, [harsūsi], est désigné dans la langue par *ḥərsīyət*.
- Pour le *bathari*, [baṭhari], on trouve chez B. Thomas *botahari*, *bautahari*.

Tous ces noms de langues ont été empruntés à l'arabe dialectal de la région (remarquons qu'ils sont au masculin en arabe et au féminin en sudarabique).

- Pour le *hobyot*, nous avons conservé le nom sudarabique ([həwbyūt]), que l'on pourrait arabiser en "*hobi*".

19.1.5. *Parenté sémitique*

Il existe plusieurs théories sur la place exacte du sudarabique. Celles qui le rapprochent particulièrement de l'araméen ou de l'akkadien le font sur des bases désormais considérées comme inexactes. Tout confirme au contraire la proximité, dans un groupe sud-sémitique, avec le sudarabique ancien, les langues sémitiques d'Ethiopie (guèze, amharique ...) et l'arabe. Voir D. Cohen 1988.

Il est désormais acquis que l'existence de consonnes glottalisées et celle de fricatives latérales sont deux traits du sémitique commun. Le premier ne s'est conservé que dans les langues éthiopiennes et sudarabiques (modernes et peut-être anciennes). Le second ne s'est conservé que dans les langues sudarabiques (modernes et peut-être anciennes).

On trouvera ci-dessous les correspondances phonétiques, illustrées d'exemples, qui mettent en évidence cette parenté.

Aussi loin dans le temps que remontent nos connaissances des populations arabes, nous trouvons la division: Arabes du Nord et Arabes du Sud; les premiers parlaient des *langues nordarabiques*, desquelles a émergé l'arabe, les autres ont inscrit dans la pierre les langues que l'on regroupe aujourd'hui sous le nom de *sudarabique ancien* ou *épigraphique*. A. F. L. Beeston (1984) propose *Sayhadic languages* pour cette famille de langues dont les premières inscriptions apparurent toutes autour du prolongement occidental du grand désert *al-rub' al-xālī*. En effet, cette bande de sable, aujourd'hui *ramlat al-sab'atayn*, était nommée *ṣayhad* au moyen-âge. La

parenté des langues sudarabiques modernes avec celles de ces inscriptions datant des royaumes antiques du Yémen (Saba', Ma'in, Qatabān, Ḥaḍramūt ...) est très probable, mais les recherches récentes montrent qu'il est exclu de voir dans les langues dites modernes le simple résultat de l'évolution de langues anciennes attestées. Elles proviennent assurément de langues parlées dans la région à l'époque des inscriptions, dont il n'a pas été trouvé de traces écrites.

Les dernières inscriptions sudarabiques précèdent de quelques années les premières observations linguistiques des savants arabes. Les données que l'on peut tirer de ces auteurs ne concernent qu'exceptionnellement—et sur des points mineurs—le sudarabique; ils accordent leur attention au *himyarite*, terme qui désignait des dialectes arabes très singuliers. Ces dialectes ont sans doute coexisté avec le sudarabique et laissé des traces dans des dialectes arabes yéménites jusqu'à nos jours.

19.2. Phonétique et phonologie

Le SAM considéré dans son ensemble présente des traits originaux au sein de la famille sémitique. Nous exposons brièvement d'une part ce qui constitue cette originalité et d'autre part ce qui caractérise chacune des langues à l'intérieur de la famille SAM. Pour une comparaison sémitique générale, voire chamito-sémitique, nous renvoyons à D. Cohen (1988, en particulier 33–35 et 11–14).

Une présentation des traits communs de la phonétique et de la phonologie des langues SAM se fait sur la base de la comparaison avec l'arabe et plus généralement avec les langues sud-(ouest-)sémitiques. C'est pourquoi, nous commencerons par donner un tableau général des correspondances consonantiques dans les langues sud-sémitiques, pour proposer ensuite le tableau du système des consonnes du SAM, avec les précisions nécessaires sur certaines articulations caractéristiques.

19.2.1. Consonnes

19.2.1.1. Correspondances

En ce qui concerne les consonnes, la comparaison entre les mots des différentes langues sémitiques met en évidence des correspondances très régulières.

Le cas trivial est celui des phonèmes qui apparaissent inchangés dans les mots homologues des diverses langues.

Ainsi reconstitue-t-on facilement le mot **baʕl*, ‘maître’, dans la langue antique qui est à l’origine des langues sémitiques, à partir de la série des mots suivants signifiant ‘maître’:

akkadien *bēlu*, ougaritique et phénicien *bʕl*,³ hébreu *baʕal*,
araméen *baʕalā*, arabe *baʕl*, guèze *bāʕl*, tigrigna *baʕal*, tigré *bāʕl*,
amharique *bal*, sabéen *bʕl*, mehri *bēʕl*, *bēl*, *bāl*, harsusi *bāl*,
bathari *baʕl*, hobyot et jibbali *baʕal*, soqotri *baʕl*.

Tableau 19-1. Correspondances consonantiques en sémitique méridional

sémitique reconstruit	mehri	jibbali	soqotri	sudarabique épigraphique	éthiopien classique	arabe
interdentales						
*θ	θ	θ	t	θ	s	θ
*ð	ð	ð	d	ð	z	ð
*ʈ	θ, ð	θ	ʈ	θ	ʂ	ð
latérales						
*ś	ś	ś	ś	s ²	š	š
*ṣ	ṣ	ṣ	ṣ	ḍ	ḍ	ḍ
sifflantes						
*s	s	s	s	s ³	s	s
chuintantes						
*š	h	š	š/ȝ/h	s ¹	s	s
occlusives vélaires						
*ḳ	ḳ	ḳ	ḳ	ḳ	ḳ	q
fricatives post-vélaires						
*ȝ	ȝ	ȝ	ȝ	ȝ	‘	ȝ

En outre, *se présentent identiques* dans toutes ces langues les consonnes:

labiales	*f, *b, *m	fric. post-vélaire	*x
dentales	*t, *d, *ʔ, *n, *l, *r	pharyngales	*h, *‘
sifflantes	*z, *ś	laryngales	*; *h
occl. vélaires	*k, *g	semi-consonnes	*w, *y

3. Sans voyelles car les écritures ougaritique, phénicienne et sabéenne (sudarabique ancienne) ne notent que les consonnes.

La comparaison et la reconstruction sont aussi possibles entre des mots présentant des différences importantes mais montrant des correspondances régulières.

Ainsi reconnaît-on la correspondance entre *d*, *s*, *t*, *ḏ*, *š*, *ʿ*, résultats divers de l'évolution du même phonème sémitique **š*, dans les mots suivants:

‘ennemi’: akkadien *šerru*, araméen *ʾār*;

‘être hostile’: hébreu *šāraʾ*, guèze *ʾadrara*, sabéen *qrr*;

‘faire souffrir’: ougaritique *šrʾ*, arabe *ḡarra*, M et H *šər(r)*, J *šerr*;

‘frapper’: S *šer*;

comme étant celle qu'on voit aussi dans de très nombreuses séries, dont:

akkadien *raxāšu*, ougaritique *rḥš*, hébreu *rāḥaš*, araméen *rḥʿ*, arabe *raḥaḡ*, guèze *raḥaḡa*, sabéen *rḥḏ*, M et H *rəḥāš*, J *raḥāš*, S *rāḥaš*, etc... pour ‘laver’.

Dans les correspondances du tableau 19-1, nous ne faisons figurer que le sudarabique et les autres langues sémitiques méridionales.

REMARQUES sur le tableau 19-1:

Le mehri, le jibbali, le soqotri et l'arabe présentés ici sont plutôt les proto-langues que tel ou tel dialecte. Ainsi, le harsusi, le bathari et le hobyot occupent-ils implicitement dans ce tableau la même colonne que celle du mehri. Pour l'éthiopien, on considère ici le système archaïque dont témoigne l'écriture du guèze et non la prononciation traditionnelle.

Le phonème occlusif post-palatal (*g*) ou palatal sonore (A.P.I. *j*) ou encore chuintant prépalatal sonore (A.P.I. *ʒ*, ici *ʒ*) des diverses langues, est uniformément noté *g* dans ce tableau.

On constate une concordance des consonnes presque totale, les seules discordances étant les suivantes:

- la sifflante **š* s'est confondue avec *s* en arabe et en éthiopien. Il lui correspond *h* presque toujours en mehri et *h*, *y* ou *š* en soqotri; le symbole *y* représente le phonème palatal murmuré, souvent interprété comme la séquence *y+h* ou *h+y*, alors qu'il s'agit de l'articulation de *y* avec une vibration particulière (“murmure”) des cordes vocales. Dans certains cas c'est un *y* simple qui est le résultat de l'évolution: MQn *yittīt* ‘six’, ML *yəfəwwəl* ‘ventres’, pluriel de *həfəl* (en MQn ce pluriel est *fūl*: *š* > *y* > *h* > *∅*).
- Les interdentes ont été confondues en soqotri (et dans certains dialectes mehri) avec les occlusives dentales: *ḏ* > *d*, *θ* > *t*, *θ* > *t*; mais en éthiopien c'est avec les sifflantes qu'a eu lieu la confusion.

- Les fricatives vélaires x et y ne sont pas absentes du soqotri, contrairement à ce qui est généralement affirmé; ce n'est que dans certains dialectes qu'elles se sont confondues avec les pharyngales (ce qui s'est aussi produit en éthiopien, mais seulement pour $y > 'y$);
- de même la pharyngale ' ne manque que dans certains dialectes mehri.

Le tableau des correspondances met aussi en évidence la remarquable unité des langues SAM; il en ressort immédiatement l'identité des consonnes du sudarabique moderne commun, c'est-à-dire du *proto-SAM* à partir duquel ont évolué les diverses langues.

Le trait le plus saillant de ces langues est la présence de trois sifflantes /s/ /ʃ/ /sʰ/: seul le sudarabique a conservé la distinction entre ces trois phonèmes du sémitique commun. Le sudarabique épigraphique distinguait dans l'écriture trois lettres, ḥ, ʃ, ʃʰ, correspondant à trois phonèmes distincts, /s¹/, /s²/, /s³/, dont le détail de l'articulation nous échappe encore (Beeston 1984: 8–10).

19.2.1.2. *Fricatives latérales*

En SAM le phonème /s/ et son homologue emphatique /ʃ/ (d'ordinaire notée $\dot{z}$ par commodité et tradition) ont une articulation alvéolaire fricative latérale⁴ (A.P.I. ʃ) que l'on peut sommairement décrire ainsi: la langue est dans la position de l'articulation de [t], partout en contact avec les alvéoles à l'exception d'un étroit passage latéral (en général à droite) par où l'air s'échappe avec un bruit de friction amplifié au contact de la paroi intérieure de la joue et du coin de la bouche très légèrement rétracté. Notons que son homologue arabe, ṣ , connaît et a connu diverses réalisations, la plus anciennement décrite (Sibawayhi, 8ème siècle, 1889: 453) étant une latérale.

19.2.1.3. *Emphase*

L'autre trait remarquable est l'articulation des phonèmes emphatiques: les trois séries, sourde, sonore, emphatique, du sémitique (et même du chamito-sémitique) se retrouvent en SAM avec la configuration normale; mais au lieu de la pharyngalisation de l'arabe, que l'on prend à tort comme référence en général, c'est une glottalisation qui est ici la marque de l'emphase. La réalisation dominante est celle d'une glottalisée éjective: la pression de l'air dans la cavité buccale provient non du soufflet des poumons mais du piston constitué par le larynx hermétiquement clos se déplaçant vers le haut. Ce type

4. Articulation proche du ʃ du gallois ou du tʃ de certaines langues amérindiennes.

articulatoire, courant dans les langues du monde, ne se trouve aujourd'hui en sémitique qu'en éthiopien⁵ et en SAM (et dans quelques rares dialectes arabes marginaux). Ajoutons que dans le dialecte mehri de Qishn (Lonnet et Simeone-Senelle 1983), l'occlusion laryngale est relâchée provoquant une laryngalisation (*creaky voice*) qu'il est difficile de prime abord de distinguer du voisement ordinaire. Certains dialectes soqotri connaissent aussi un affaiblissement de l'emphase.

19.2.1.4. Système phonologique

Le système consonantique du SAM commun se présente donc ainsi:

Tableau 19-2. Système consonantique du sudarabique moderne commun

	sourde	sonore	glottalisée	nasale	vibrante	semi-voyelle
labiales	f	b		m		
interdentales	θ	ð	θ̤			
dento-alvéolaires	t	d	t̤	n	r	
latérales	ś	l	ś̤			
sifflantes	s	z	ś̤			
chuintante(s)	š		(š̤)			
palatale						y
vélaires (occlusives)	k	g	k̤			labio-vélaire w
post-vélaires (fricatives)	x	ɣ				
pharyngales	ħ	ʕ				
laryngales	h	ʔ				

On reconnaît là le système sémitique commun, *augmenté d'un seul phonème*, la chuintante glottalisée š̤.

5. Nous entendons par *éthiopien* la famille des langues sémitiques d'Ethiopie ("éthiosémitique" de certains auteurs).

19.2.1.5. *Le phonème chuintant glottalisé*

Il a été placé entre parenthèses pour signifier que son appartenance au SAM commun n'est pas certaine. Toutes les langues contemporaines présentent le ʃ , mais on ne trouve guère de mots en ʃ homologues dans toutes les langues.

SQa ʃaʃa' 'boire une petite gorgée (dans la main)' n'a pas d'homologue en *M, J, H*.

ML, H ʃərōm 'gifler' n'a pas d'homologue en *J, S*.

J kʃʃɔl, ML H kəʃáwl 'briser' n'a pas d'homologue en *S*. (Pour ʃ , voir §§ 19.2.1.6. et 19.3.2.1.4.)

L'articulation ʃ se rencontre dans trois sortes de mots.

a) Ceux où elle provient de la palatalisation de k^6 (parallèle à celle de k , voir ci-dessous), c'est le cas le plus fréquent:

SHr ʃédhər 'marmite' (*MQn kādər - H kédər*) - *SHr*

məʃəher/məʃéri/mákehər 'étable' - *J minʃérɔt* 'majeur' (doigt)

(*MQn* et *H mənḵīrōt*); *mīʃɔrfɔt/moḵóruḫ, moʃóruḫ - MQn ʃab-ʃāb* 'porcelaines' (cauris) (arabe *qibqib*).

b) Ceux où elle résulte d'une évolution particulière de ʃ . Ainsi le mot 'doigt' comporte-t-il un ʃ dans:

MQn haʃbá/haʃōba; MQnB ʃəbá/ʃōba; MJb hīʃəbá, īʃəbá/īʃōba;

ML ʃəbá/ʃāba' - H hīʃəbá/hīʃōba; H(ML) eʃəbá' - B(ML) haʃ-bá' - H haʃbá/həʃāba' - JE(HL) ʔəʃbá'.

Mais dans d'autres dialectes c'est la forme sémitique en ʃ qui apparaît:

J īʃbá'/eʃē' - H ṭəʃba'/ṭəʃāba'.

comme en soqotri:

SQb ʔəʃba'/əʃābə'.

D'autre part le mot mehri *ML ɣəʃ, JE(ML) ɣéʃ*, 'avalier à grandes lampées', est peut-être à rapprocher de l'arabe *ɣaʃʃa* 's'étouffer en avalant'.

c) Ceux où elle résulte d'une évolution particulière de $\text{ʃ}(\text{ʒ})$. Ainsi:

ML ʃəfdēt, H ʃəfdáyṭ 'grenouille' (arabe *ḡufda'a'*: racine *ʒfd*).

Enfin, quelques cas ne peuvent s'expliquer par k , ʃ ou $\text{ʃ}(\text{ʒ})$, tels que:

'coude' *MQn ʃaffī - H ʃīfē' - B(ML) ʃəffáyḥ - H ʃəffáy - J(ML)*

ʃəfēf (racine non attestée en soqotri); 'gros bâton' *H təkəʃɔ'*

(mais *J təkəʃɔ' - ML təkəʃē*); 'arc' *ML mənʃəbēt - H nəʃībət* (mais

J mənṭɔt).

6. La palatalisation de k en jibbali du centre (celui de *JL*) aboutit à ʃ , parallèle à ʃ qui résulte de k .

19.2.1.6. Palatalisation de *k* et *g*

La palatalisation de *k* (et de *ḳ*) est un phénomène très marqué dans les langues SAM. Il affectait déjà (dans une faible mesure) le proto-SAM puisqu'un mot sémitique en *k* est attesté avec *š* dans tout le SAM:

'foie': MQn, ML et B(ML) *šəbdīt* - H *šəbdēt* - J *šəbdét* - S *šəbdeh*
- H *šəbdét*.

En jibbali du centre le son palatalisé est *ṣ̣* (décrit dans JL et Johnstone 1984) qui se distingue donc de *š* au point que dans ce dialecte il constitue un phonème indépendant. Une paire minimale importante est celle qui oppose *-ṣ̣* (pronom suffixe 2fs) à *-š* (pr. suff. 3ms). Notons aussi *yšérək* 'il fait' :: *yšérək* 'il grimpe'. (Pour *ṣ̣*, voir aussi § 19.3.2.1.4.)

Synchroniquement les alternances *k/š(ṣ̣)* sont très nombreuses dans toutes les langues mais particulièrement productives en jibbali. En soqotri on trouve aussi des degrés de palatalisation différents: *c* (occlusive palatale) et *ç* (éjective) dans certains dialectes:

SQb *ycōtəb* 'il écrit' (*ktb*); *dça'r* 'de la maison' (< *di-ka'r*).

On observe aussi une palatalisation de *g* en soqotri et en jibbali:

SHr *tenēžəh* sorte de pamplemousse, SQa *yažəh* 'femme';

J *yég/yóži/yég*, dim. *yížég* 'homme' - *yažét* 'grande fille' (cf. SQa ci-dessus). (Pour *ž*, voir aussi § 19.3.2.1.4.)

19.2.2. Voyelles

19.2.2.1. Timbres

Pour ce qui est des voyelles, la seule caractéristique générale est, par rapport au sémitique commun, un élargissement de la gamme des timbres: y figurent toujours les voyelles *a*, *i*, *u*, *o*, *e*, *ε*, *ə*. Ces timbres s'organisent dans des systèmes assez variés, présentant ou non une opposition brève/longue, une nasalisation, etc. Ils seront précisés plus loin lors de la présentation de chaque langue. Il n'est pas possible encore de proposer un système de voyelles pour le proto-SAM, qui serait intermédiaire entre le système sémitique (*a u i* - *ā ū ī*) et les systèmes actuels; il faut en particulier déterminer si les voyelles *o* et *e* étaient ou non en place avant la séparation des langues.

19.2.2.2. Correspondances

Dans les langues SAM, lorsque les lexèmes se correspondent, et dérivent donc a priori du même mot sud-sémitique, il arrive, dans un nombre de cas limité seulement, que les timbres des voyelles soient voisins ou identiques:

'genou' *bark* en M, B, S, H, et *berk* en H, J. (< **bark*);

‘chèvres’ *MQn hēráwn, H həwərūn, H həywurōn, B(ML) ʾāʾrān, J ʾerūn, S ʾrəhən* (< *ʾarōn < *ʾarān).

Le plus souvent, les voyelles sont différentes, résultant de règles historiques d'évolution qui produisent des correspondances régulières:

ML həfhōs H ʾefhós J efhés ‘cuire à l'eau’ (< *ʾhafhas);

ML həylōt H aylōt J aylét ‘se tromper’ (< *ʾhaylat).

On aperçoit à travers trois des quatre exemples précédents une évolution phonétique changeant *a* en *o* dans certaines conditions, variant selon les langues; dans le sémitique méridional, seul le SAM est dans ce cas.

Malgré l'impression de régularité que peuvent donner les exemples précédents, on constate que les voyelles s'organisent dans des systèmes phonologiques structurellement très différents; il en résulte des correspondances très complexes encore obscurcies par une harmonisation vocalique partielle dans certaines langues et des discordances morphologiques entre les langues portant précisément sur les voyelles, la structure syllabique et l'accent.

‘tendon d'Achille, talon’ *ML mədrēm J mədrúm SQb médrəhem;*

‘sourcil’ *MQn gəfōn J gəfn SHo gēfən;*

‘rein’ *ML kəlyīt J kuLét SQb kəlšyət.* (Pour *L*, voir § 19.3.2.1.3.)

On ne s'étonnera donc pas de ne pas trouver ici de tableau de correspondances vocaliques pour le SAM. Il est cependant possible de donner des séries de correspondances entre les schèmes vocaliques des formes nominales et verbales qui constituent la base du lexique.

a) noms

*sud sém. *CaCaC*

M CəCēC

J C(ə)CəC, CəCəC

*sud sém. *CaCC*

M CaCC, CōCəC, CēCəC

J CəC(ə)C, CəCC, CəCC

b) suffixe nominal de féminin

*sud sém. *-at*

M -ōt, -ēt, -ət, -t

J -ōt, -ét, -ət, -t

S -əh, -eh, -əh, ...

c) verbes

*sud sém. *CaCaCa/yaCCuCu/yaCCuC*

M CəCūC/yəCūCəC/l-əCCēC

J C(ə)CəC/yəCəC/yəCCəC

S CəCəC,

CCəC/yəCəC/l-iCCēC

*sud sém. *CaCiCa/yaCCaCu/yaCCaC*

M CīCəC/yəCCōC/l-əCCōC

J CēCəC/yəCēCəC/yəCCəC

S CēCəC/yəCēCəC/l-iCCəC, &c.

19.2.2.3. Rôle des voyelles dans la morphologie

a) Dans les verbes, les voyelles participent aux schèmes de la dérivation verbale, à l'opposition de genre, de nombre, de personne et de diathèse, selon le fonctionnement régulier des langues sémitiques. Mais une particularité est l'existence de certaines alternances vocaliques pour les oppositions de genre et de nombre.

- genre (masculin/féminin)

'tu écris' *MQn təkūtəb\tekītəb*; *J tkōtəb\tkītəb*; *SQa tkōtəb\tkētəb*

'que tu écrives' *J tōktəb\tkītib*.

- nombre (singulier/pluriel)

ML kətūb 'il a écrit'/'*kətawb* 'ils ont écrit'.

J yfēḏōr 'il tremble de peur'/'*yfēḏēr* 'ils tremblent de peur'.

SQa ykōtəb 'il écrit'/'*ykētəb* 'ils écrivent'; *ktōb* 'il a écrit'/'*kteb* 'ils ont écrit'.

b) Dans les noms, de même, on remarque le rôle particulier de l'alternance vocalique pour les oppositions de genre (S et J seulement) et de nombre essentiellement dans les quadrilatères:

- genre (masculin/féminin)

SQa tāḥrər\ṭāḥrer 'chèvre sauvage', *śībəb\śībīb* 'vieux', *xálxal\ xálxel* 'grisonnant' (poil); *S(JL) šəzrhar\šəzrhīr* 'jaune', *kérkam\kérkim* 'jaune'.

J šḥamúm\šḥamím 'brun de peau', *ṣaḥbōb\ṣaḥbéb* 'brun clair', *šəzrōr\šəzréb* 'jaune', *ṣəfrōr\ṣəfrér* 'jaune', *xəzgōg\ xazgéḡ* 'fin' (tissu).

- nombre (singulier/pluriel):

H ʾardīb\ʾardōb 'nuque' - *MQn bəṣṣén\bəṣṣón* 'articulation de la hanche' - *J məṣṣḥérér\məṣṣḥérōr* 'os du tibia' - *SQb nāḥrər\ nāḥrur* 'nez'.

19.2.2.4. Variation morpho-phonologique

Des phénomènes de variation affectent les noms et les verbes: quand ils sont construits avec un pronom suffixe personnel, on observe des modifications de timbre vocalique, de durée, d'accent, de structure syllabique.

MQn ʾabšōš & ʾabšēšīä 'les lèvres' & 'mes lèvres'; *MQnB ɸaṣawl & ɸaṣaleh* 'il a brisé' & 'il l'a brisé'; *ḥōrəm & ḥārʾmi* 'chemin' & 'mon chemin';

H ʾəlbədtō & ʾəlbədtāyhəm 'elles deux ont frappé' & 'elles deux les ont frappés'; *bəxāš & bəxṣēni* 'cela a fait mal' & 'cela m'a

fait mal'; *yəmōt* & *yémtəh* 'il a maltraité' & 'il l'a maltraité';
yərōb & *yərbéni* 'il a reconnu' & 'il m'a reconnu';
B nōka & *neke'ēn* 'il est venu' & 'il est venu à nous'; *B(ML) hātər*
 & *hštri* 'chevreau' & 'mon chevreau';
J šúyəl & *ššyəlš* & *šúylən* & *šuylək* 'travail' & 'son travail' &
 'notre travail' & 'ton travail'; *yéxtər* & *yəxtərsən* 'qu'il retire
 du feu' & 'qu'il les retire du feu'; *škəl* & *éḳíləš* 'il a fait
 coucher' & 'il l'a fait coucher'; *effōšəl* & *effīšílšhum* 'il
 embarrassa' & 'il les embarrassa'.

19.2.3. Syllabe - accent

19.2.3.1. Mehri et langues voisines

On se bornera ici à indiquer que le système le plus simple, et sans doute celui qui a le moins évolué, est celui du mehri.

En règle générale, la *syllabe* est de type CV(C) ou C $\bar{V}$, à l'initiale (C)CV(C) ou (C)C $\bar{V}$, à la finale CV(C(C)) ou C $\bar{V}$ (C).

L'*accent* tombe sur la (dernière) syllabe longue {CVC(C), C $\bar{V}$ (C)} ou, si toutes les syllabes sont brèves {CV}, sur la première (CVC en fin de mot compte pour brève).

ML məlūk 'posséder'; *kālāyn* 'le soir'; *rīḳōḳ* 'fins'; *bāhi* 'par eux deux'; *yəhānsəməm* 'ils respirent'; *təhənāwdəxən* 'elles fument'; *mənḳərbəṭəwtən* 'tordues de douleur'.

Il y a quelques exceptions: *ML fənwəhi* (? < *fənwīhi*) 'devant eux deux', et le suffixe *-ən* de certaines conjugaisons est hors unité accentuable: *ML yarkə-báyən* (*yarkəbē+ən*) 'eux deux mettent sur le feu (une marmite)'; les voyelles non phonologiques ne comptent pas: *ML ənxálihəm* (= /nxalihəm/) 'sous eux'; *θəbəkəm* (< *θəbrəkəm*) 'vous êtes cassés'. Le 'compte comme une consonne, même lorsqu'il a disparu laissant pour trace une influence sur la voyelle (ouverture, allongement): *yənákam* (< *yənák'əm*) 'ils viennent'. Il en est de même pour '.

Au contact de deux mots, une voyelle de disjonction se présente pour rétablir la structure syllabique menacée par le contact des consonnes:

MQn ḥamš-ə-ttīyəs 'je veux que tu la manges' [-š, -s: pr. suff. 2fs, 3fs].

Cependant certains groupes consonantiques sont admis:

MQn ḥamk ttah 'je veux que tu le manges' [-k, -h: pr. suff. 2ms, 3ms].

De plus, de tels groupements admissibles se produisent dans les mots, par chute d'un ə :

MQn dekkk 'ton coq' (*dekk-k*) - *MQnB fitkkōt* (*f*syllabique) 'elle s'est détachée, libérée'.

Le harsusi et le bathari gardent l'accentuation du mehri malgré l'abrégement de certaines voyelles:

H šəlabəd (< **šəlēbəd*) 'il a essayé de tirer (au fusil)'; *təxām* (< **təxaym*) 'tu veux'; *käbəlēt* (< **käbəlēt*) 'sud'.

Le harsusi admet la syllabe CVC en position non finale: *yeķerōbhem* 'il s'approche d'eux'; de même que le bathari: *a'ayāntha* 'ses (m.) yeux'.

Les langues autres que le mehri occidental ont, en plus, la possibilité d'avoir une voyelle à l'initiale de mot:

J efkäh 'briser en deux'; *énuf* 'lui-même' - *ML amōdēh* 'flatter'; *aḡāyg* 'l'homme' - *H aḡšōm* 'rafraîchir'; *aḡāh* 'son frère' - *B exuft* 'la plante du pied'.

Il s'agit en général d'un préfixe de dérivation verbale ou d'un article défini.

19.2.3.2. Le système jibbali (jibbali du centre)

Il a connu une évolution particulière. Le contraste $\bar{V} :: \check{V}$ est perdu et l'accent n'est pas de culmination: un mot peut avoir plusieurs syllabes également proéminentes: *ənkəzbəzət* 'elle a soupiré'.

Lorsque le mot mehri apparenté existe, avec un schème analogue, la comparaison montre que, lorsque le mot jibbali n'a qu'un accent, celui-ci est situé sur la syllabe homologue de la syllabe accentuée du mot mehri.

MLJ: nīdəm\nīdəm 'se repentir'; *həndūm\lendīm* 'aider'\faire repentir'; *nāddəm\nūtdəm* 'regretter'; *šəndūm\šəndīm* 'revenir'; *nədmīt\nədmēt* 'regret'.

Mais on peut pousser la comparaison plus loin: dans une série de mots de même schème, le nombre d'accents varie selon l'environnement consonantique; la dernière syllabe accentuée correspond alors régulièrement à celle qui porte l'accent en mehri.

MLJ: həsūf\hsəf 'il a humilié'; *həgūs\həgəs* 'il a pensé' (< **CaCaCa*); *kəfēn\kfēn* 'linceul'; *ḡəlēt\ḡālēt* 'faute' (< **CaCaC*).

On peut en conclure que le proto-jibbali avait la même accentuation que le mehri et que l'énergie accentuelle a eu tendance à se répartir vers le début du mot.

De ce point de vue, le jibbali de l'est est conservateur:

ML *kəláwkəl*, JE *kələbkəl*, J *kələbkəl* ‘rafales de vent’
 (< **kalākil*); ML *məḵhayrī*, JE *məṣḥayré*, J *məṣḥérér* ‘tibia’.

Une exception importante à la correspondance formelle entre le mehri et le jibbali est celle du subjonctif des verbes simples de type transitif (< sémitique **fa’ala*, par opposition à **fa’ila*):

ML *kəṭūb/ykūṭəb/yəktēb*, J *ktəb/ykətəb/yəktəb* ‘écrire’.

Mais divers indices, dont la 1^{ère} personne du pluriel *nəktēb* et l’impératif J *ktəb* conforme au M *ktēb*, permettent de supposer que le subjonctif en J ait eu secondairement une transformation accentuelle particulière.

19.2.3.3. Le système soqotri

Le soqotri a poussé cette tendance plus loin: la syllabe initialement accentuée a perdu son accent qui est remonté vers le début du mot, en même temps que l’opposition de durée était abolie. Comparer:

‘morceaux de bois’ M *ḥayrūb*, J *ḥəṣəb*, SQa *təyrəb*;
 ‘cinq’ (+ nom masculin) MQn *xmoh*, H *xəmmóh* J *xōš*, SQa
xōmāy;
 ‘écorce d’arbre’ MQn *kalīfót*, J *kiLifót*, SQa *kályēfəh*;
 ‘enfant’ H *məbərə*’ J *əmbéré*’ SQa *məbrəhe*’;
 ‘feuille’ MQnB *ṣəfāṭ*, J *zyiLfót*, H *ṣəyelfót*, SQa *ṣəyləfəh*;
 ‘il vole (oiseau)’ MQn *yifərū*, J *yəfiré*, SQa *yífror*.

Dans le cas où une voyelle longue perd son accent et sa durée, il s’est développé un dédoublement de cette voyelle au moyen de *h*: J *kaṭmim*, S *kāṭməhim* ‘beurre’. C’est ce qu’on appelle le ‘*h* parasite’, voir § 19.3.3.1.2.

19.2.4. Gémination consonantique⁷

19.2.4.1. Existence de la gémination

En comparant à l’arabe, on constate l’absence de la gémination comme morphème de dérivation verbale. Pour opposer ‘mettre en garde’ :: ‘prendre garde’, par exemple, l’arabe utilise une gémination à valeur morphologique: arabe *ḥaḍira* :: *ḥaḍḍara*. Le SAM ne présente, dans les formes équivalentes, que des consonnes simples:

M *ḥəḍūr* :: *ḥəḍər* - J *ḥəḍər* :: *oḥḍūr* - S *ḥəḍər* :: *ḥəḍər*.

Et ceci même dans des emprunts à l’arabe: *mu/a’allim* ‘professeur’ est devenu J *m’alm* et MQn *mōləm*.

Il en a été tiré, à tort, la généralisation selon laquelle *il n’y aurait pas de gémination* en SAM. Or il suffit, pour prouver le contraire, de présenter les

7. Voir, pour le jibbali, Johnstone 1980.

quelques faits suivants qui concernent les langues autres que le soqotri (voir REMARQUE le concernant).

La gémination est attestée partout mais de façon variable selon les langues et les dialectes; ainsi pour le mot ‘coude’:

MQn *šaffi/šaffyūtən*; *MMf* *šafi/šafōwyət*; *ML* *šaffáy/šafōf*, *šə-fāwwət*, *šəfūtən* - *B* *šəffáyh* - *H* *šəffáy/šəffāyət* - *H* *šīfē/šəfiy-ōtə*, *šəffōtə* - *J* *šéfēf/šəffətə*.

Comparer aussi: *MQn* *xaff* mais *ML* *xaf* ‘pied’.

19.2.4.2. Types et variation

Il existe en SAM deux causes de gémination phonétique.

a) Soit la présence d’une racine de type $C_2 = C_3$ ou $C_3 = C_4$ (gémination d’origine radicale):

MQn *šəllūt* ‘elle a pris’; *ML* *kəṭəbbūt* ‘poupée en bois’; *kəṭəf-fāwtən* ‘petites plumes’ - *J* *dəkkēt* ‘banc’; *xazəggūn*, ‘très fins’ (vêtements) - *B* *dəssīt* ‘pierre taillée’ - *H* *heddōt* ‘berceau’; *yemellēt* ‘nuages amoncelés’ - *H* *yəllōt* ‘brouillard’;

on a même $C_4 = C_5$ dans *MQn* *šxūllūt* ‘elle s’est assise’ (*šxwll*).

b) Soit le contact (et l’assimilation) de la consonne du morphème infixé *-t-* avec une consonne radicale homorganique ou quasi-homorganique (gémination d’origine morphologique):

MQn *‘āššad* (‘*-t-šd*’) ‘être inquiet’; *fəttāwn* (‘*f-t-tn*’) ‘essayer de se souvenir’; *kāssər* (‘*k-t-sr*’) ‘être humilié’; *ML* *kāššəl* ‘être cassé’ (‘*k-t-šl*’); *kāššər* (‘*k-t-sr*’) ‘être dépouillé’; *xāzzəl* ‘plaisanter’ (‘*x-t-zl*’); *xāššəl* ‘être percé’ (‘*x-t-šl*’).

Dans le dialecte de *ML*, *t* s’assimile avec: *d*, *t*, *š*, *ṣ*, *z*, *s*, *ṣ*, *θ*, *ḏ*, *θ*.

Le traitement des formes comportant une gémination (de l’une et l’autre origine) est très original en SAM. Il suffit d’observer le paradigme de verbes tels que:

J *terr/yətrér/yōttər* ‘tirer’ - *MQn* *haḵrāwr/yhaḵrāwr*, *f*.
haḵarrāwt/lhāḵḵar, participe actif *məhāḵḵar* ‘partir’ - *ML*
rāttəšəm/yərtəšəys/yərtəššəm (pluriel) ‘être en foule’,

pour constater que, dans la variation morphologique, la gémination consonnantique ne suit pas toujours la même consonne; selon les langues et les dialectes, les situations sont variables et complexes; nous ne donnerons ci-dessous qu’un aperçu de ce transfert de gémination.

REMARQUES

- Même sans transfert de gémination, la différence avec l’arabe est remarquable: si à l’accompli arabe, par exemple, *tamma* ‘être terminé’ correspond

le mehri (*MQn*) *təmm*, à l'inaccompli *yatimmu* correspond *MQn* *yitəməūm* (et le subjonctif *lətmēm*) sans gémiation.

- La réorganisation syllabique en présence d'un suffixe peut faire apparaître une gémiation phonétique: *J* *əhdéd* 'il a menacé', *əhəddəš* 'il l'a menacé'.

19.2.4.3. Transfert de gémiation

19.2.4.3.1. VERBE ET PARTICIPE

La gémiation porte, au gré de la morphologie, selon les formes dérivées, les personnes ..., sur la deuxième *ou la première* consonne radicale (troisième ou deuxième pour les quadrilitères), ou non-radical dans le cas du *-t*-infixé:

J *kerr/yəkrér/yókkur* 'sauter' - *MQn* *šhəzáwz/yšhəzáwz/lšəhəz*
'suivre à la trace' - *J* *axləl* 'laisser passer la pluie' (toit), partici-
cipe actif *míxxəl; aḥəðnín/yḥəðnín/yḥəððən* 'regarder fixe-
ment'.

Avec *-t*-:

H <šll> *šáttəl/yəštəlól/yəštəl* 'transhumer' - *MQnB* <fk> *fəttək*
'se détacher' (f. *fikkōt*).

REMARQUES

- Le transfert d'une gémiation d'origine morphologique provoque une gémiation sur les racines non géménées. Ainsi, *MQn* *kaššáwr* (< *k-t-šr*) 'rétrécir (au lavage)' a pour féminin *kašarrūt* dans lequel l'infixation de *-t*-n'est plus apparente, l'apparence étant celle de **kšrr*.

- Une autre conséquence du transfert de gémiation s'observe dans certaines homonymies telles que *J* *kóttəb* 's'écrire les uns aux autres' <*k-t-tb*> et *kóttəb* 'être réticent' <*k-t-bb*> (transfert).

19.2.4.3.2. NOM

Il existe des noms comportant une consonne géminée; leur pluriel présente dans certains cas un transfert de gémiation.

a) Gémiation d'origine radicale:

ML *kəwwét* 'force'; *MQn* *sənnórt/sənnórər* 'chatte' - *H*
kəttəbót/kəttəbáb 'poupée' (taillée dans le bois ou l'os).

b) Gémiation d'origine morphologique (schème sémitique **CaCCāC*):

MQn *bəhḥšr* 'marin'.

c) Par ailleurs, certains noms déverbaux à préfixe comportent, par transfert, une gémiation différente de celle de la racine:

J *middék* 'percuteur de fusil' (*dkk*), *múxxud* 'épingle à cheveux'
(*xdd*).

d) La gémiation apparaît aussi dans une série de pluriels:

ML *kāṣar/kaṣáwwar* ‘pièce au premier étage’; *kōban/kabúwwan*
‘bouse sèche’ - H *kaḥōr/kaḥáḥḥar* ‘chamelle suitee’.

e) En outre, il existe, en mehri, hobyot et harsusi, quelques occurrences de gémiation difficilement explicables:

ML *ḥannūk, ḥḥennek* ‘palais et gencive’ (J *ḥónúk, MQn ḥank*).
H *ḥamméni* ‘huit’ (MQn *tamōni, J ḥōni*).

19.2.4.4. Remarques concernant la gémiation en jibbali

Le jibbali présente aussi une gémiation de la première consonne du mot lorsque lui est préposée une voyelle, ce qui se produit dans trois circonstances pour les verbes et une pour les noms, à condition que cette première consonne soit dans la liste suivante: $C_1 = \acute{s}, \acute{s}, \tilde{s}, f, t, k, y$ (dans les verbes, occasionnellement *d, θ, g, z, ɣ, x, h, h*, et exceptionnellement *d, ž*).

a) Les verbes intensifs-conatifs (correspondant aux formes II et III de l’arabe) sont de type *eCCóCeC*, les causatifs (forme IV de l’arabe) sont de type *(a)CCeC*, et les quadrilitères de type *(e)CeCCéC*. Tous développent, tantôt à l’accompli, tantôt à l’inaccompli, une gémiation de C_1 :

- *eCCóCeC*:

eśśókər/iśśókərən/iśśókər ‘plisser des yeux au soleil’;
eθθúkəl/iθθúkələn/yθθókəl ‘charger’; *axlé/íxxóle/yáxlé* ‘consacrer’;

- *(a)CCéC*:

etlék/ittélók/yétlək ‘marcher derrière une bête attachée par la patte’;
eklé/ikkóle/yékle ‘amener à la maison’ (mariée, ses animaux);
aḥrér/iḥḥrér/yéḥḥar ‘parquer les chevreaux’.

Remarquer que pour ce dernier verbe la gémiation *ḥḥ* est, à l’indicatif, due à la voyelle préfixée, et, au subjonctif, au transfert de *rx*:

- *(e)CCeCCéC*:

ettértér ‘mener rudement’.

b) Les noms déterminés par l’article (constitué par une voyelle):

ɔ-xxóbz ‘le pain’; *a-xxaf* ‘la sole du chameau’, mais *ɔ-(x)xófét* ‘la fenêtre’; *ɔ-(x)xórf* ‘les pluies de la mousson’; *e-ffá’m* ‘le pied’, mais *e-fénús* ‘la lampe’; *e-kkób* ‘le chien’, mais *e-(k)kēr* ‘le cheikh’; *e-tté* ‘la viande’; *i-yyúm* ‘le jour’, ‘le soleil’; *e-ssáḥart* ‘la sorcière’; *e-śśá’b* ‘la vallée, le cours d’eau’; *e-ššō* ‘le dos’, mais *e-(š)śúm* ‘le nom’; *e-ššírś* ‘le ventre’.

Un phénomène de *gémiation compensatoire* apparaît lorsque, pour des raisons de contexte la première consonne radicale est élidée (c'est le cas de *b* à l'intervocalique), la deuxième radicale est alors géminée:

J ebhím/iḥhúm (< **ibḥúm*) 'ne pas se décider'; *ebšél/yēššəl*
'cuisiner'.

Dans le cas où *C*₂ est une consonne qui s'élide, c'est *C*₁ qui est géminée:

J ezbét/ízzíṣt 'donner du lait'.

19.2.4.5. Remarques concernant la gémination en soqotri

D'une façon générale, la gémination est quasi absente en soqotri; elle est en général réduite: *SQa* 'əṣ, f. 'əṣḥ/yəṣḥ/*l-iṣáṣ* 'avoir peur'; et *S sfəh* à côté de *MQn səffīt* 'cheveu'.

- Gémination d'origine radicale (dans les noms exclusivement):

SQaB 'ərebḇəḥ, *SQb* 'ərebəḇəḥ 'cuisse'; *S kəṣṣəḥ* 'cheveux coupés'.

- Gémination d'origine morphologique:

S hódde (< *ḥ-t-dr*) 'prendre garde'; *lōttəm* 's'envelopper le visage';
házzək 'plaisanter'.

- Gémination d'origine phonétique. La voix murmurée (cf. supra) provoque un allongement consonantique qui peut parfois aller jusqu'à une gémination:

SQa léš^hen, *SQaB léšš^hen* 'langue'.

Pour le mot *SQa mēšhe*, *S(ML) mēsse* 'pluie', on peut aussi penser au résultat de l'assimilation de *l*, la racine étant <lsww> (*MJb malsé*).

Notons enfin que l'apparition de la gémination est parfois limitée à certains dialectes.

19.3. Particularités des langues

Nous présenterons ici les trois langues les mieux connues, le mehri, le jibbali et le soqotri. Les autres seront évoquées incidemment, le harsusi et le bathari étant phonétiquement très voisins du mehri, le hobyot étant, selon ses dialectes, proche du mehri de Jadib ou du jibbali.

19.3.1. Mehri

Le mehri est la langue qui a le moins divergé du proto-SAM, la langue parlée avant la séparation en plusieurs langues: elle n'a connu ni les transformations phonologiques du soqotri et du jibbali, ni les développements morphologiques et syntaxiques du soqotri.

19.3.1.1. Consonnes⁸ (voir Lonnet et Simeone-Senelle 1983)

19.3.1.1.1. SYSTÈME COMPLET

Le système consonantique est le système complet du sudarabique, c'est-à-dire celui du sémitique, comme le montre le tableau comparatif (tableau 19-1); une seule exception, mais très caractéristique du mehri (et, dans une moindre mesure, du soqotri), est l'évolution de *š en *h*.

MQn hemm 'nom' (J *šum*, S *šem*);

MQn ħroh 'tête' (J *rēš*, *SHr rī/rīhi/ʿrʾəš*);

MQn bəhēl 'être cuit' (J *béšəl*, S *béhel*);

dans un cas le *h* passe même à *ħ*: *mẓrāħ* 'dent' (arabe *ǧīrs*) où *š* > *h* > *ħ*.

Il reste quelques *š* non passés à *h*, tels que:

ML šəṭāwf, J *šəṭəf* 'blesser légèrement'; *ML ləšēn*, J *elšén*, *SHo*

līšin 'langue'.

19.3.1.1.2. INTERDENTALES

Dans quelques dialectes, essentiellement citadins, s'est produite la coalescence des dentales et interdentes. C'est le cas bien connu du parler de Qishn où 'sein', 'cuisse', 'sur' se disent *tōdi*, *fxād*, *ṭār*, mais ailleurs *θōdi*, *fxāð*, *θār*.

19.3.1.1.3. FAIBLESSE DE '

La fricative pharyngale sonore (le 'ayn de l'arabe) manque partiellement, parfois presque totalement, dans certains dialectes, en particulier à Qishn, Ṣaqr, et dans le dialecte du Dhofar, décrit dans *ML*.

MQnMJb: *ākāb\akāb* 'oiseaux'; *təṛā\ṭəṛā* 'elle garde le trou-peau', *šīrē\šīrē* 'nombril'.

Il arrive que la même racine se manifeste chez le même locuteur avec et sans 'ayn: *lāθīt/lī'áθəṭən* 'scinque', sorte de lézard.

19.3.1.1.4. RÉTROFLEXES

Apparemment plus fréquentes à l'ouest qu'à l'est sont les réalisations rétroflexes des groupes *r* + dentale.

MQn mərðəbōn [ɾɖ] sorte de jarre; *MMf rēzəm/hārzōm* [ɾʂ] 'dessus du pied', *ḵornēt* [ɾŋ]/*ḵarān* 'utérus'.

Notons que la durée de la voyelle qui précède le [ɾ] est allongée, et que ce [ɾ] rétroflexe peut avoir des réalisations faibles, jusqu'à disparaître, ce qui rappelle la situation de certains dialectes de l'anglais.

8. Nous avons, en notant systématiquement *g*, neutralisé les diverses réalisations dialectales : ce *g* représente [j] de *MQn*, le *g* de *ML*, etc.

19.3.1.1.5. VÉLARISATION DE *l*

Inversement, on observe, à l'est et pas à l'ouest, la vélarisation de *l*. Le contexte est celui d'une fermeture de syllabe non finale:

ML həwbūs (MQn həlbūs) 'habiller qqn'; *kəwθēt (MQn kəltēt)*
'histoire, conte'.

Et ceci même quand la voyelle qui précède est une voyelle épenthétique:

ML əwbōn (MQn ləbōn) 'blanc'.

Dans le cas où la syllabe est accentuée, on peut avoir chute de *l* et allongement de la voyelle: *ML sēmək* 'je suis sauf' (< *səlmək*).

Cependant, la vélarisation ne se produit pas quand *l* est suivi d'une pharyngale ou d'une laryngale.

19.3.1.1.6. DÉVELOPPEMENT DE 'r'

Le *ħ* et le *h* initiaux, non étymologiques, sont souvent le développement d'une occlusive glottale [ʔ].

'père' (*ʔb*) *MQn H həyb*; *H həb*; *J ʔy* (*b > Ø*);

'mère' (*ʔm*) *MQn hāmē*; *H hām*; *H hāmē*; *J ʔm*;

'chemin' (*ʔm*) *MQn hawrēm*; *H wōrēm*; *H ḥōrēm*; *J ʔm*;

'femmes' (*ʔn*) *MJb ḥaynēθ*; *H ḥáyneθ*; *H ḥānīθ*; *J ʔnēθ*;

et peut-être parfois la trace d'un article qui ne subsiste fonctionnellement qu'en mehri du Dhofar:

'tête' (*rš < rʕ* et *h < š*) *MQn ḥroh*; *H ḥérih*; *H ḥarēh*; *J rēš*.

19.3.1.1.7. GLOTTALISATION À LA PAUSE

Dans certains dialectes, on observe une fermeture de la glotte après la dernière voyelle lorsqu'elle est longue et suivie d'une consonne. Il en résulte une articulation éjective de cette consonne:

MJb dəmēɣ > [dəmēɣʔ] 'cerveau', *MQnB bīr > [bīrʔ]* 'puits'.

Lorsqu'il s'agit d'une nasale, elle se réduit à un souffle nasal (sourd), la nasalité pouvant affecter la voyelle.

19.3.1.2. Voyelles

Le système vocalique comprend deux ou trois voyelles brèves (selon les dialectes) *a*, *ə/ʌ*, (*ε*), six voyelles longues: *ā*, *ē*, *ē*, *ī*, *ō/ɔ*, *ū* et deux diphtongues *ay*, *aw*. Du point de vue diachronique un certain nombre de régularités apparaissent. En comparant le verbe de base *kəṭūb/yəkūtəb/yəktēb* 'écrire' qui comporte des consonnes 'neutres' aux verbes *nəḥāg/yənūḥəg/yənḥāg* 'danser, jouer', *nəḥāwr/yənūḥər/yənḥār* 'vider un chargeur', on constate l'influence des consonnes telles que pharyngales et vélaires (*ħ*, *x*, *ɣ*) d'une part, et des emphatiques (glottalisées) d'autre part. Les mêmes phénomènes se produi-

sont dans les noms: comparer *ḵəffēt* ‘panier’ et *ḵəṣṣāt* ‘frange’, *ḥərūz* ‘acacia’ et *ḵəzāwb* ‘fauve’ (non identifié).

Nous donnons ici quelques exemples de paires minimales pour des voyelles de faible contraste (*MQn*):

hətt ‘six’ :: *hətt* ‘donne!’

kəbkēb ‘entrée’ :: *kəbkīb* ‘étoile’

tawmər ‘elle dit’ :: *tōmər* ‘dattes’

fām ‘jambe’ :: *fawm* ‘jambes’

ktūb ‘il a écrit’ :: *ktōb* ‘livre’

yāyḍəl ‘il a porté’ :: *yādəl* ‘porter’

tləbyən ‘tu (m.) remplis’ :: *tləbyən* ‘tu (f.) remplis’.

19.3.1.3. Diphtongues

Nous renvoyons aux exemples de mehri, de hobyot et de harsusi du présent article: l’abondance des diphtongues *aw* et *ay* y est manifeste par rapport à leur rareté en jibbali et en soqotri. Elles sont équivalentes à une voyelle longue et majoritairement conditionnées par la présence d’une pharyngale, d’une vélaire ou d’une emphatique; mais certaines diphtongues ne peuvent s’expliquer par ce conditionnement.

19.3.2. Jibbali

Le jibbali se caractérise, phonétiquement, par la transformation de son accentuation (voir § 19.2.3.2), par la vocalisation des bilabiales *b* et *m*, l’occlusion de la bilabiale *w* ou son amuïssement, la réalisation fricative de *l*, la perte de l’opposition de quantité vocalique du proto-SAM, compensée par l’abondance des timbres vocaliques. Ces phénomènes empêchent l’identification directe de la racine. Ainsi, la plante *izyet* est-elle *wəblīt* en mehri (même racine *wbl*).

En outre, le jibbali du centre, non celui de l’est, présente des sifflantes particulières (à côté de *š*, *ś*, *s*) fortement labialisées: *ṣ̣*, *ṣ̣̣* et *ž*.

Les exemples ci-dessous sont uniquement pris en jibbali du centre, celui de *JL*, sauf indication contraire, *JE*, pour le jibbali de l’est.

19.3.2.1. Consonnes

19.3.2.1.1. VOCALISATION DE *b* ET *m*

b et *m*, en position intervocalique se vocalisent, produisant les seules voyelles longues de la langue (nasalisées dans le cas de *m*):

erḥīt, f. de *erḥīm* ‘beau’ (*MQn rḥaym*, *rḥáymət*);
yōri (*ML ḡābōrī*) ‘djinn’s; *ḥēr* (*ML ḥáybər*) ‘se refroidir’;
 ou produisant des hiatus: *JE ḡiér* (*MQn ḡābēr*) ‘dauphin’; *JL ḡiél* (*MQn ḡābél*) ‘montagne’. Dans ce dernier cas le pluriel a été formé sur le singulier réduit: *ḡélātə* (mais *MQn ḡābəlīn*).

Il arrive qu’un mot, dont la racine contient deux *b*, soit réalisé avec un ou zéro *b*, selon les contextes: **sabab* > **sābeb* > *siēb* ‘cause’, et avec suffixe -*Vk*: *essiēk*.

19.3.2.1.2. EVOLUTION DE *w*

Le *w* radical passe à *b* ou à une voyelle, selon les contextes, et n’est normalement jamais réalisé *w*:

šābhéd ‘être livré à soi-même’ (*wḥd*), *egbéz* ‘rendre licite’ (*gwz*);
ēkt ‘temps’ (*wkt*), *éléd* (*wld*) ‘enfants’.

La voyelle peut être longue: *əštōḥ* (*šwḥ*) ‘pleurer amèrement’.

19.3.2.1.3. *L*: /FRICATIF

Dans des contextes palatalisants, *l* peut avoir une variante fricative: *L* (notée *ž* dans *JL*). Elle est sonore et ne peut être confondue avec la sourde *s* ni avec l’éjective *š* (notée *ẓ̌* dans *JL*): *ḡiLṣl* (*ML ḡīlṣl*) ‘nourriture cuite’.

19.3.2.1.4. SIFFLANTES LABIALISÉES *š*, *ṣ* ET *ž*

Ces articulations sont “réalisées avec approximativement la même position de la langue que pour *š*, mais il n’y a pas de contact entre le haut de la langue et les alvéoles. L’air est expulsé au-dessus de la langue et les lèvres sont simultanément arrondies en une moue” (*JL*: xiv).

- *š* (voir § 19.2.1.6) a pris une indépendance phonématique par rapport à *ṣ̌*. Rappelons la paire minimale fondamentale des pronoms suffixes -*š* 3ms :: -*ṣ̌* 2fs (confondus en *JE*); et ajoutons *šābdún* ‘courageux’ :: *šābdún* ‘à plat ventre’. C’est souvent une simple variante palatalisée de *k*: *šīrs/ekrēs* ‘ventre’.

Mais elle peut s’opposer à *k*: *ešḥéf* ‘perdre’ :: *ekḥéf* ‘faire affronter’.

- *ṣ* est son homologue emphatique (voir § 19.2.1.5). Elle s’oppose à *ḵ*:

ṣeff ‘coudre grossièrement’ :: *ḵeff* ‘se détourner’.

Mais elle en est le plus souvent une variante:

fōḵḥ/fūṣḥi ‘une/deux moitié(s)’.

- *ž* (voir § 19.2.1.6) est une simple variante palatalisée de *g*: *žénúzt/généz* ‘cadavre’.

19.3.2.2. Voyelles

L’opposition de quantité originelle a disparu et les timbres sont nombreux: *a*, *ε*, *e*, *i*, *ə*, *u*, *o*, *ɔ*.

De nouvelles voyelles longues sont produites par l'intégration de *b*, *w* et *m* (voir § 19.3.2.1.1–2). Dans ce dernier cas, il y a en outre nasalisation. L'article défini s'amalgame à une voyelle initiale (précédée de ') et l'allonge.

Nous présentons ici quelques paires minimales particulièrement révélatrices.

rayáz 'devenir doux' :: *rəyáz* 'doux'
rěš 'tête' :: *rěš* 'têtes'
erdé 'rejeter' :: *erdé* 'abandonner'
ḥáləl 'licéité' :: *ḥálúl* 'laxatif'
ʿoz 'chèvre' :: *ʿoz* 'la chèvre'
ʾayá 'frère' :: *ʾāyá* 'le frère'
ʾiróz 'riz' :: *ʾiróz* 'le riz'
lun 'sorte' :: *lūn* 'blanc'
essólm 'il sauva' :: *əssólm* 'que je sauve'
éléd 'enfants' :: *íléd* 'qu'il ait des enfants', ...

Quant à *a* et *ε* ils sont en distribution complémentaire: *a* est toujours en contexte vélaire ou pharyngal; de même *u*, *o* et *ɔ* sont vraisemblablement les variantes de deux phonèmes, bien que les notations de Johnstone laissent subsister un doute: comparer *ub* 'coeur', *ʾob* 'non', *ʿob* 'porte', *ʿob* 'la porte'.

La disparition de *w* a entraîné la perte des diphtongues *Vw*. Il y a quelques rares diphtongues *Vy*: *ʾiy* 'père'. Plus fréquent est le hiatus: *yšīūn* 'il croit' (*ʾmn*); *modōi* 'étables' (*dwm*); *idīās* (*dyb*) 'il l'a punie'.

19.3.3. Soqotri

Le relatif isolement géographique du soqotri a permis des évolutions linguistiques particulières qui en font une langue très originale dans le groupe sudarabique et dans la famille sémitique. Dans la brève présentation qui suit, on trouvera ces traits originaux et d'autres caractéristiques résultant d'évolutions plus répandues dans les langues sémitiques.

19.3.3.1. Consonnes

19.3.3.1.1. COALESCENCE DE PHONÈMES

a) Dans tous les dialectes connus, les interdentales *θ*, *ð*, *θ* se sont confondues avec les occlusives dentales *t*, *d*, *t*. Ainsi 'sang' (racine sudarabique *ḏrw/y*): *dər^h*, *dōr*, *dōr*,... selon les dialectes soqotri, mais *ḏor* en jibbali. De même *S ʾarb* 'morceau de bois', *trih* 'deux', *M H B H J* mêmes mots, mais avec *θ* et *θ*.

b) Inversement, nous sommes désormais en mesure d'affirmer que ce n'est que dans certains dialectes que les fricatives vélaires γ et x se sont confondues avec les pharyngales 'et $ħ$:

SQa tɣəd, SHo, Qb tɣ'əd 'tu vas'; *SQa xɔmɔɣ, Qb ɣɔymɔħ*
'cinq'.

19.3.3.1.2. LE *h* PARASITE

Son caractère "parasite" apparaît quand on compare certains mots soqotri à leurs homologues d'autres langues sudarabiques ou d'autres langues sémitiques. Le *h* intercalé perturbe la correspondance entre les consonnes des racines. Ainsi, le mot 'oreille' (racine ∂n) se présente-t-il sous les formes: akkadien *uzn*, ougaritique *ʾudn*, hébreu *ʾozen*, araméen *ʾudnā*, arabe *ʾudn*, éthiopien *əzn*, mehri *ħayðēn*, jibbali *ʾidēn*, harsusi *ħayðēn*, hobyot *ħáyðən*, et soqotri *ʾidehen*. De même: *S dəkəħən* ($\partial k n$) 'menton', 'barbe'; *ʾThóntən* (γn) 'yeux'; *ʾırhez* (γz) 'riz', etc.

Ce *h* n'a aucune valeur morphologique, il est le résultat d'une évolution phonétique des voyelles longues dont une reconstitution est proposée par Leslau (LS: 22). Voir aussi Bittner (1913: 5 n.), Prætorius (1908: 708–13) et Rhodokanakis (1915: 12–56).

19.3.3.1.3. VOIX MURMURÉE

Parallèlement au "*h* parasite", phonème de plein statut, existe une "voix murmurée", apparemment réalisée par une ouverture partielle des cordes vocales (en vibration ou non selon le contexte). Ce type phonatoire se propage sur un, deux phonèmes, rarement plus. Nous le notons par un ^h, placé arbitrairement après une consonne murmurée, alors que la voyelle suivante et/ou la voyelle précédente peuvent être aussi murmurées:

SQa kār^h 'gorge', *lēs^hən* 'langue', *t^háʿeləh* 'ver'.

C'est peut-être un développement de ce murmure qui a permis l'émergence du *h* "parasite". Dans *gém^hel/gemīli/géməħəL* 'chameau', le singulier comporte un murmure, le pluriel un *h* parasite, l'un et l'autre absents du duel.

Inversement, le *-h* du féminin singulier nominal et verbal (accompli 3ème personne) est parfois réduit à un murmure de la voyelle précédente ou même à zéro.

Ce murmure provoque sur // l'apparition d'une variante légèrement fricative (ici notée *L*). Dans un dialecte (Qalansiya) la friction est telle que la réalisation de ce phonème converge avec celle de /ʒ/ (souvent déglottalisée en soqotri). Ainsi pour le mot 'côte', nous avons relevé: *žal*ʹ, *žəl*ʹ, ... et, à Qalansiya, *žáz*ʹ.

19.3.3.1.4. ARTICULATION DE /ʔ/

Elle n'est fricative pharyngale sonore qu'entre deux voyelles. Ailleurs elle est, dans certains dialectes occidentaux, occlusive ou plutôt affriquée: occlusive pharyngale à détente fricative pharyngale. Cette occlusion est, à la première écoute, difficilement discernable d'une occlusion glottale. Quant à la friction, elle est normalement sonore mais en fin de mot elle est dévoisée comme toute consonne. Il s'ensuit la perception de [ʔʰ#] (dans les autres dialectes il y a généralement un simple [ʔ#]).

19.3.3.1.5. PALATALISATION (voir § 19.2.1.6)

La palatalisation contextuelle de /k/ connaît des degrés divers selon les dialectes. Elle peut aboutir à [c] qui ne semble pas avoir un statut de phonème dans les dialectes que nous avons observés; quant à /g/ ([ʃ], [g]) sa réalisation avancée [ʒ] (A.P.I. [ʒ]) a, au moins dans un dialecte, un statut phonologique à faible rendement d'opposition. Nous n'avons relevé qu'une paire minimale: *SQaB máḥže* 'lobe de l'oreille' :: *máḥge* 'enclos'.

19.3.3.1.6. AFFAIBLISSEMENT DE L'EMPHASE

Plusieurs dialectes ont une articulation faiblement éjective pour les occlusives (*t*, *k*) et quasi-sonore pour les fricatives (*z*, *ṣ*, *ṣ̣*). Lorsque *k* se palatalise (voir § 19.2.1.5), l'affrication quasi-sonore fait converger la consonne avec *g*: *SQa gárbaḵ/gərbēgi/gərbīḥ* ou *gərbēbeg* 'chat'. On trouve même, dans certains dialectes, la convergence *k-ž*: 'abreuer' *héžə*, dont le réfléchi est *téḵi* 'boire' (racine *šḵy*).

19.3.3.2. Voyelles

L'opposition de longueur n'apparaît qu'exceptionnellement dans les dialectes connus et avec un rendement infime. Un résidu des voyelles longues sémitiques et proto-sudarabiques est sans doute le *h* parasite (cf. ci-dessus).

Le système vocalique est riche en timbres dans tous les dialectes; certains comportent moins de voyelles labialisées que les autres, mais l'inventaire phonématique est conservé, [ɔ] et [o] étant remplacés par des réalisations [ɤ] et [u] qui restent en opposition avec les voyelles antérieures.

Ainsi *SQa* présente, pour l'inaccompli de *sed* 'mettre à chauffer sur le feu', les formes suivantes (sans préfixe, cf. Johnstone 1968): 2ms [syd], 2fs [sɪd], 2mpl [səd] qui, dans d'autres dialectes sont respectivement [sɔd], [sid], [səd].

Signalons que le timbre *u* n'apparaît pratiquement que dans le passif inaccompli (sans marque personnelle) des verbes: *SQa tūḵəṭ* 'il se fait réveiller' (actif: *itāḵəṭ* 'il réveille'). On remarque que, dans l'ensemble,

contrairement au mehri, les timbres moyens dominent largement les timbres extrêmes.

Bibliographie⁹

Abréviations

BSOAS: *Bulletin of the School of Oriental and African Studies*

JSS: *Journal of Semitic Studies*

KAWW: Kaiserliche Akademie der Wissenschaften in Wien

MAS-GELLAS: *Matériaux Arabes et Sudarabiques*. Paris: Univ. Paris-3, Groupe d'Etudes de Linguistique et de Littératures Arabes et Sudarabiques

SAE: Südarabische Expedition

SBKAWW: *Sitzungsberichte der KAWW*, phil.-hist. Klasse

ZAL: *Zeitschrift für arabische Linguistik*

ZDMG: *Zeitschrift der deutschen morgenländischen Gesellschaft*

Beeston, A. F. L. 1984. *Sabaic Grammar*. Manchester: University of Manchester (JSS Monograph 6).

Bittner, Maximilian. 1909–15. “Studien zur Laut- und Formenlehre der Mehri-Sprache in Südarabien, 1–5.” *SBKAWW* 162/5, 168/2, 172/5, 174/4, 176/1, 2, 178/3.

———. 1913–18. “Vorstudien zur Grammatik und zum Wörterbuche der Soqotri-Sprache, 1–3.” *SBKAWW*, 173/4, 186/4, 5.

———. 1916–17. “Studien zur Šḥauri-Sprache in den Bergen von Dofâr am Persischen Meerbusen, 1–4.” *SBKAWW* 179/2, 4, 5, 183/5.

Cohen, David (éd. par). 1988. *Les langues chamito-sémitiques*. Paris: C.N.R.S. (Les langues dans le monde ancien et moderne, éd. par J. Perrot, 3).

Fresnel, Fulgence: voir Lonnet 1991.

al-Hamdani (10ème siècle). 1884–91. *Al Hamdānī's Geographie der arabischen Halbinsel [Šifat Jazīrat al-‘Arab]*. Ed. par D. H. Müller. Leyden: Brill.

———. 1983. *Šifat Jazīrat al-‘Arab*. Ed. par M. b. ‘A. al-Akwa‘. San‘ā’: Markaz al-dirāsāt w-al-buḥūṭ al-yamanī.

Hein, Wilhelm. 1909. *Mehri- und Ḥaḍrami-Texte gesammelt im Jahre 1902 in Gischin von Dr. —, bearbeitet und herausgegeben von Dav. Heinr. Müller*. Vienne: Alfred Hölder (KAWW, SAE, 9).

9. Nous donnons ici une bibliographie sommaire. Pour approfondir, on se référera à Leslau (1946) et Robin (1977).

- Jahn, Alfred. 1902. *Die Mehri-Sprache in Südarabien. Texte und Wörterbuch*. Vienne: Alfred Hölder (KAWW, SAE, 3).
- . 1905. "Grammatik der Mehri-Sprache in Südarabien." *SBKAWW* 150/6.
- Johnstone, T. M. 1968. "The Non-occurrence of a *t*-Prefix in Certain Socotri Verbal Forms." *BSOAS* 31: 515–25.
- . 1970. "A Definite Article in the Modern South Arabian Languages." *BSOAS* 33: 295–307.
- . 1975a. "The Modern South Arabian Languages." *Afro-Asiatic Linguistics* 1/5: 93–121.
- . 1975b. "Contrasting Articulations in the Modern South Arabian Languages." *Hamito-Semitic*. Ed. par J. et Th. Bynon, 155–59. La Haye: Mouton (Janua Linguarum Series Practica, 200).
- . 1977. *Ḥarsūsi Lexicon and English-Ḥarsūsi Word-List*. Londres: Oxford University Press (= *HL*).
- . 1980. "Gemination in the Jibbali Language of Dhofar." *ZAL* 4: 61–71.
- . 1981. *Jibbāli Lexicon*. Londres: Oxford University Press (= *JL*).
- . 1984. "New Sibilant Phonemes in the Modern South Arabian Languages of Dhofar." *Current Progress in Afro-Asiatic Linguistics: Papers of the Third International Hamito-Semitic Congress*. Ed. par J. Bynon, 389–90. Amsterdam-Philadelphie: Benjamins (Current Issues in Linguistic Theory, 28).
- . 1986. "Mahri." *Encyclopédie de l'Islam*, nouv. éd., 6: 82–83.
- . 1987. *Mehri Lexicon and English-Mehri Word-list* (= *ML*), avec: *Index of the English Definitions in the Jibbāli Lexicon*, compiled by G. Rex Smith. Londres: SOAS.
- Leslau, Wolf. 1938. *Lexique Soqotri (Sudarabique Moderne) avec comparaisons et explications étymologiques*. Paris: Klincksieck (Coll. ling. pub. par la Soc. de Ling. de Paris, 41). (= *LS*)
- . 1946. "Modern South Arabic Languages. A Bibliography." *Bulletin of the New York Public Library* 50/8 (August): 607–33.
- Lonnet, Antoine. 1985. "The Modern South Arabian Languages in the P.D.R. of Yemen." *Proceedings of the Seminar for Arabian Studies* 15: 49–55.
- . 1991. "La découverte du sudarabique moderne: le eḥkili de Fresnel (1838)." *MAS-GELLAS* n.s. 3: 15–89.

- Lonnet, Antoine et M.-C. Simeone-Senelle. 1983. "Observations phonétiques et phonologiques sur les consonnes d'un dialecte mehri." *MAS-GELLAS* 1: 187–218.
- Matthews, Charles. 1967–70. "On the Border of the Sands." *Univ. of South Florida Language Quarterly* 6/1–2: 39–47, 6/3–4: 7–12, 7/1–2: 41–48, 7/3–4: 43–48, 8/1–2: 43–47, 8/3–4: 11–19.
- Morris, Miranda. 1983. "Some Preliminary Remarks on a Collection of Poems and Songs of the Baṭāḥirah." *Journal of Oman Studies* 6: 129–44.
- Müller, David Heinrich. 1902. *Die Mehri- und Soqotri-Sprache. I. Texte*. Vienne: Alfred Hölder (KAWW, SAE, 4).
- . 1905. *Die Mehri- und Soqotri-Sprache. II. Soqotri-Texte*. Vienne: Alfred Hölder (KAWW, SAE, 6).
- . 1907. *Die Mehri- und Soqotri-Sprache. III. Šḥauri-Texte*. Vienne: Alfred Hölder (KAWW, SAE, 7).
- Naumkin, Vitalij V. et V. Ja. Porxomovskij. 1981. *Očerki po etnolingvistike Sokotry* [Essais d'ethnolinguistique de Soqotra]. Moscou: Nauka.
- Prætorius, Franz. 1908. "Zur Frage über das parasitische *h* des Minäischen." *ZDMG* 62: 708–13.
- Rhodokanakis, Nikolaus. 1915. "Der Zweigipflige Akzent im Minäo-Sabäischen (Studien zur Lexicographie und Grammatik des Altsüdarabischen, I)." *SBKAWW*, 178/4: 12–56.
- Robin, Christian (éd. par). 1992. *L'Arabie antique de Karib'il à Mahomet. Nouvelles données sur l'histoire des Arabes grâce aux inscriptions*. Aix-en-Provence: Edisud (*Revue du monde musulman et de la Méditerranée* 61, 1991–93).
- [Robin, Christian]. 1977. "Langues et dialectes modernes de l'Arabie du sud." *Bibliographie générale systématique, Corpus des inscriptions et des antiquités sud-arabes*, chap. 18, 89–99. Louvain.
- Sibawayhi (8ème siècle). 1881–89. *Le livre de —, Traité de grammaire arabe*. Paris. Ed. par H. Derenbourg. Repr. Hildesheim - New York: Olms, 1970.
- Simeone-Senelle, Marie-Claude. 1991a. "Récents développements des recherches sur les langues sudarabiques modernes." *Proceedings of the International Hamito-Semitic Congress, 1987*, Ed. par Hans G. Mukarovskij, vol. 2, 321–37 Vienne (Beiträge zur Afrikanistik, 41).
- . 1991b. "Notes sur le premier vocabulaire soqotri: le Mémoire de Wellsted (1835). Première partie." *MAS-GELLAS* n.s. 3: 91–135.

- . 1992. “Notes sur le premier vocabulaire soqotri: le Mémoire de Wellsted (1835). Deuxième partie.” *MAS-GELLAS* n.s. 4: 13–82.
- Simeone-Senelle, Marie-Claude et A. Lonnet. 1986. “Lexique des noms des parties du corps dans les langues sudarabiques modernes. Première partie: la tête.” *MAS-GELLAS* 3: 259–304.
- . 1988–89. “Lexique des noms des parties du corps dans les langues sudarabiques modernes. Deuxième partie: les membres.” *MAS-GELLAS* n.s. 2: 191–252.
- . 1991. “Lexique Soqotri: les noms des parties du corps.” *Semitic Studies in Honor of Wolf Leslau on the Occasion of His Eighty-fifth Birthday, November 14th, 1991*. Ed. par Alan S. Kaye, vol. 2, 1443–87. Wiesbaden: Harrassowitz.
- . 1992. “Compléments à *Lexique Soqotri: les noms des parties du corps*.” *MAS-GELLAS* n.s. 4: 85–108.
- Thomas, Bertram. 1929. “Among some Unknown Tribes of South Arabia.” *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 59: 97–111.
- . 1937. “Four Strange Tongues from South Arabia: the Hadara Group.” *Proceedings of the British Academy* 23: 231–331.
- Wagner, Ewald. 1953. *Syntax der Mehri-Sprache unter Berücksichtigung auch der anderen neusüdarabischen Sprachen*. Berlin: Akademie Verlag (Deutsche Akademie der Wissenschaften zu Berlin, Institut für Orientalforschung, Veröffentlichung 13).
- Wellsted, James R.: voir Simeone-Senelle 1991b, 1992.

Abréviations et Conventions de lecture

Langues

- a) Les exemples sont tirés des corpus relevés par nous sur le terrain, sauf:
- B*: bathari relevé par M. Morris, sauf, si précisé (*ML*), par Johnstone.
- H(ML)*: hobyot du Dhofar relevé par Johnstone dans le *Mehri Lexicon*.
- H*: harsusi, dans le *Harsūsi Lexicon* (Johnstone 1977).
- J*: jibbali du centre, dans le *Jibbāli Lexicon* (Johnstone 1981).
- JE*: jibbali de l’est relevé par Johnstone.
- LS*: soqotri (Müller), compilé dans *Lexique Soqotri* (Leslau 1938).
- ML*: mehri du Dhofar, dans le *Mehri Lexicon* (Johnstone 1987).

b) Nos propres données sont introduites par l'abréviation du nom de la langue suivi du lieu d'enquête (voir la carte):

MQn: mehri de Qishn.

MQnB: dialecte bédouin de la région de Qishn.

MMf: mehri de Muḥayfif (région d'al-Ghayḍah).

MJb: mehri de Jādib.

S: soqotri.

SHo: soqotri de Ḥadibo.

SHr: soqotri de la montagne Ḥagher (dialecte bédouin).

SQa: soqotri de Qalansiya.

SQaB: soqotri des bédouins de la région de Qalansiya.

SQb: soqotri de Qadhub.

H: hobyot de la région de Ḥawf.

Autres abréviations

SAM sudarabique moderne; sém. sémitique; pr. pronom; suff. suffixe;
m. masculin; f. féminin; 2ms = 2ème pers. masculin singulier, etc...

Autres conventions de lecture

La morphologie des noms est présentée ainsi:

singulier/duel/pluriel ou: singulier/pluriel;

et celle des verbes:

accompli/inaccompli indicatif/inaccompli subjonctif (à la 3ème
personne du masculin singulier).

Chaha (Gurage) Phonology

Wolf Leslau

University of California, Los Angeles

Chaha is a Gurage (Guragué) dialect.¹ It belongs to the branch of the Semitic language family of which the others members are Ge'ez (at present the language of the liturgy), Tigre, and Tigrinya in the north; Amharic, Argobba, Harari, Gafat, and Gurage in the south.

Gurage has at least twelve dialects, some of them having variants. There are three distinct groups of this cluster: West Gurage, which includes Chaha, Endegeñ, Inōr (or Ennemor), Eža, and Gyeto; East Gurage, which includes Selṭi, Wolane, Ulbarag (or Urbarag), and Zway; North Gurage, with Soddo as its only representative; and Muher, Masqan, and Gogot, the positions of which still remain to be investigated.

Note that Chaha is called *čäha* by the speakers, and *čäha* or *čəha* by the Amharas.

East Gurage is linguistically related to Harari, and Soddo (or Kəstanəñña) is related to Gafat (Leslau 1950b). As for West Gurage, even though it has South Ethiopian features and a few notable North Ethiopian features, it still has no particularly close connection with any specific South Ethiopian language. For the position of Muher, Masqan, and Gogot, see above.

Hetzron (1972, 1977) suggested another classification, but see Leslau 1969; 1979, 1: xi–xii.

The region of Gurage is situated southwest of Addis Ababa, the capital of Ethiopia, at a distance of about 120 miles. It is bordered on the north by the river Awash, on the east by Lake Zway, and on the south and west by the river Omo.²

1. ABBREVIATIONS: Amh., Amharic; Ar., Arabic; C, consonant; C., Chaha; E., Eža; Ed., Endegeñ; En., Ennemor; G., Ge'ez; Go., Gogot; Gt., Gyeto; Had., Hadiya; Har., Harari; Hebr., Hebrew; Kam., Kambata; M., Muher; Ms., Masqan; Qab., Qabenna; Sid., Sidamo; So., Soddo; Te., Tigre; Tña., Tigrinya; V, vowel; W., Wolane; Z., Zway. A lexeme *not* preceded by C. refers to Chaha.

2. For more details, see Leslau 1979, 1: ixff. For a description of the Western part of Gurage, see Shack 1966.

Most normally, the point of departure is a Chaha feature in relation to the feature of the other Gurage dialects. Occasionally a phonetic feature is mentioned in the various Gurage dialects with the exclusion of Chaha.

20.1. Labials and labialization

The labials are: *b m f w*, allophone *p*; rounded *b^w m^w f^w*, allophone *p^w*.

20.1.1. *b*

b corresponds to Semitic and Ethiopic *b*. Example: C. *bāna* (for *l* : *n*, see § 20.7.1.2) 'eat', Ar. *balī'a*, G. *bāl'a*, Amh. *bälla*.

Minimal pairs *b:b^w* in *bər* 'silver' : *b^wər* 'main'; *bārä* 'leaf of the *äsät*-plant that is torn' : *b^wārä* 'cow that has a white spot on the forehead'; *bəṭər* 'kind of bush' : *b^wəṭər* 'eighth of a loaf of bread'.

In intervocalic or final position *b* may become *w*. Examples: C. *āwawt* 'roundworm' (Go. *ābabut*); C. *āwä* 'dew' (G. *həbo*); C. *səwä* 'animal fat' (from *säbba* 'be fat'); C. *šəwät* 'gray hair' (from *šəbät*); C. *nəzəw* 'flexible' (from **nəzəb*, *ləzəb*).

b became *y* in C. *əngäyä* 'kind of berry' (Gt. *əng'aba*).

-*äbV*- became *o* in C. *wəg'o* 'violent rain with wind' (Go. *wäg'äbä*); C. *aso* 'salt' (S. *asäbo*).

-*əbə*-, -*əb* became *u* in *sur* 'broken' (from *səbəṣ*, *səwer*); *tu* 'breast' (from *təb*, *təw*).

Alternation between C. *b* and Gurage *m* occurs in C. *qaribo* 'kind of beer' and M. *qarimo*.

Alternation between C. *b* and *f* occurs in C. *bəṭərq barä*³ 'feel excitement' and M. *fəṭərq beä*; C. *anqəbäqäbä* and *anqəfäqäfä* 'cut down branches'.

The labial *b* is occasionally inserted after *m*: C. *ambər* 'cabbage' (W. *aməl*); C. *qambisa* and *qämis* 'shirt' (*qämis* borrowed from Arabic).

For devoicing of originally geminated *b*, *b^w* becoming *p*, *p^w*, see § 20.10.1; § 20.10.2.

20.1.2. *b^w*

b^w of the various Gurage dialects may become *w* in Chaha: thus, C. *wäz* 'slave' (En. *b^wäz*); C. *wəyä* 'go down to a lower place' (Ms. *b^wəyä*); *däwä* 'relatives' (E. *däb^wä*).

3. The final -*m* of the Chaha perfect (*barä-m*) will be omitted.

Intervocalic: C. *a-wāka* ‘make dough’ (from *a-b^wāka*); *yä-wānet* ‘of the husband’s sister’ (for *yä-b^wānet*); *yä-wəya* ‘of the *b^wəya*-liquid’.

b^w instead of *w* by pseudo-correction (see § 20.17) in *b^wər* ‘main’ instead of *wər*.

Note C. *b^wə*- as against *bu* of the other Gurage dialects: C. *b^wəko* ‘dough’, E. *buko*.

20.1.3. *f*

f corresponds to *f* of Semitic and Ethiopic: C. *fāta* ‘untie, loosen’, Ar. *fataḥa* ‘open’, G. *fāṭha*.

Minimal pairs *f*: *f^w* in C. *āf* ‘mouth’: *āf^w* ‘bird’; *fākārā* ‘multiply’: *f^wākārā* ‘boast’.

f occasionally nasalizes the preceding *a*: thus, *āfrax^a* ‘pregnant cattle’; *āfwat* ‘odor’; *zāf^a* ‘afterbirth of cattle’.

For the alternation *f*: *b*, see § 20.1.1.

Note that Chaha *f^wä* becomes *fo*, and C. *f^wə* becomes *fu* in the other Gurage dialects. Thus, C. *f^wägägä anṭä* ‘cut off something very close’, but Gt. *fögägä anṭä*; C. *f^wəgəg* ‘last day of the full moon’, but E. En. *fugəg*.

20.1.4. *m*

m corresponds to *m* of Semitic and Ethiopic: *mānṭä* ‘strip off’, Ar. *malāṭa* ‘shave’, G. *mālätä* ‘strip off’. (for *l*: *n*, see § 20.7.1.2).

Minimal pairs of *m*: *m^w* in C. *mar* ‘beeswax’: *m^war* ‘share’; *mena* ‘action’: *m^wena* ‘maternal uncle’; *əməṛ* ‘stone’: *əmə^wər* ‘strength’.

m in dialects other than Chaha may become *w*: thus, *w* of Ed. *əwān* ‘louse’ goes back to *m* of M. *qəmal*, C. *qəmar* (for *l*: *ɾ*, see § 20.7.3); Ed. *žəwān* ‘red core of the trunk’, but C. *žəmar*; Ed. *qəwāččä* ‘be bashful’, but C. *qəməččä*.

Final *-äm* of Ethiopic is reduced to *-o* in Chaha: e.g., G. *māskārām*, name of the 1st month, > C. *māsxäro*.

Final *-m* is lost in *axu* ‘you’ (pl.) for *axum*.

Note that Chaha *m^wä* becomes *mo* and *m^wə* becomes *mu* in the other Gurage dialects. Thus, C. *m^wänä* ‘bladder’, but Ed. *monä*; C. *m^wəst* ‘ill-mannered’, but En. *must*.

20.1.5. *w*

w corresponds to *w* of Semitic and Ethiopic: e.g., C. *wāṭäqä* ‘fall’, Ar. *wadaqa* ‘fall in drops’ (for *ṭ* from *d*, see § 20.14.1.5), Amh. *wäddäqä* ‘fall’.

20.1.5.1. *w from consonants*

$w < b$: thus, C. *äwa* ‘dew’ (G. *həbo*); C. *äwawt* ‘roundworm’ (Go. *äbabut*).

$w < b^w$: C. *wäz* ‘slave’ (En. *b^wäz*); C. *wəyä* ‘go down to a lower place’ (Ms. *b^wəyä*); C. *däwä* ‘relatives’ (E. *däb^wä*).

$w < m^w$: thus, C. *wämaka* ‘proverb’ (En. *m^wam^waka*); C. *wətətənyät* ‘upper part of the rolled-up leaf of the *äsät*-plant’ (Gt. *mutətənyät < m^wətətənyät*).

For $w < m$ by dissimilation, see § 20.15.3.

$w < k^w$ in C. *näwšašä* ‘the young of a goat’ (Ed. *nokšašä* from *näk^wšašä*), and probably C. *wašwašinyät* and *k^wäšk^wäšiyä* ‘kind of bush’.

$w < g^w$ in C. *wərawər* ‘throat’ and *g^wärärä*.

$w < h$ in C. *wəsača* ‘day when the *äsät*-plant is scraped’, from Kambata *hasačča*.

20.1.5.2. *Contractions involving w*

wä > o: C. *wägät* and *ogät* ‘public discussion’.

-äw > o: C. *tona* ‘strength’ (Gt. *ṭäwna*).

aw > o: C. *od* ‘threshing field’ (En. *awd*); C. *ozat* ‘flour boiled in water’ (Gt. *awzat*).

-äwä > o: C. *boxe* ‘well’, from *bäwäxe*; *bogäret* ‘well’, from *bäwägäret*; *yorčä* ‘of the front leg’ from *yäwärčä*.

-äwə > o: C. *yokäränä* ‘one who participates in the monthly gathering’, from *yäwəkäränä*; *yosa* ‘of the bread made of the *äsät*-plant’, from *yäwəsa*.

awä > a, ɔ: e.g., C. *aṭa, ɔṭa* ‘bring out’, from *awäṭa*; *andä, ɔndä* ‘lower’ (from **awändä*, M. *awärrädä*).

20.2. Labialization

While the labiovelars and the rounded labials are phonemes in Gurage and in the other Ethiopian languages outside of Tigre and Harari, they are also phonetically conditioned.

In the impersonal of Gurage of the type ‘one says’ a labializable consonant is labialized. The labializable consonants are the labials $b > b^w$, $f > f^w$, $m > m^w$, $p > p^w$, and the velars $g > g^w$, $k > k^w$, $x > x^w$, $q > q^w$ (see Leslau 1967). Thus, *b^wänem* (from *bäna-m* ‘eat’); *čäf^wärim* (from *čäfärä-m* ‘put into the mouth’); *m^wänem* (from *mäna-m* ‘fill’); *säp^wärim* (from *säpärä-m* ‘break’); *šäg^wärim* (from *šägärä-m* ‘change’); *məsäk^wärim* (from *məsäkärä-m* ‘bear witness’); *x^wärim* (from *xarä-m* ‘know’); *qäṭärim* (from *qäṭärä-m* ‘kill’).

20.3. Velars

The velars are: *g*, *k*, *q*; spirant *x*; palatalized *g'*, *k'*, *x'*; *q'*; labiovelars *g^w*, *k^w*, *x^w*, *q^w*.

20.3.1. Velar *g*

g corresponds to Arabic *ğ* (*g*), Ethiopic *g*: e.g., C. *gamera* ‘camel’, Ar. *ğamal*, G. *gämäl*.

In Arabic loanwords, C. *g* corresponds to Ar. *ğ* (*g*), *ġ*: e.g., C. *ləg^wam* ‘bridle’, Ar. *luğām*; C. *gaz* ‘battle’, Ar. *ġaz^w*.

Minimal pairs *g* : *g'* in *gārā* ‘buttocks’ : *g'ārā* ‘leaf of the *āsāt*-plant that is not torn’; *gātārā* ‘put to sleep’ : *g'ātārā* ‘be lenient toward someone’.

Minimal pairs *g* : *g^w* in *gaz* ‘raid’ : *g^waz* ‘caravan’; *gəbər* ‘utensil’ : *g^wəbər* ‘hunchbacked’; *gār* ‘mild’ : *g^wār* ‘harvest’.

For the originally geminated *g*, *g'*, *g^w* becoming *k*, *k'*, *k^w*, see § 20.10.4; § 20.10.5; § 20.10.6.

For *g* becoming palatalized *g'*, see § 20.9.

For labialization of *g* into *g^w*, see § 20.2.

C. *g^w* alternates with *w* in *g^wārāro* ‘throat’ and *wərawər*.

C. *gu-* (from *g^wə-*) may become *w* in the other Gurage dialects: thus, C. *gudella* (from *g^wədella*) ‘strap around the tail of the horse’ as against M. So. *wəddəlla*.

C. *g^wä* may become *go* in the other dialects: thus, C. *g^wägä* ‘well, pit’, Ed. *goğä*.

20.3.2. *k*

k corresponds to *k* of Semitic and Ethiopic: C. *kāfana* ‘shroud’, Ar. *kafana* ‘envelop in a shroud’, Te. *kāfna*.

Minimal pairs of *k* : *k'* in C. *wāka* ‘beam’ : *wāk'a* ‘loan of a cow’.

Minimal pairs in *k* : *k^w* in C. *tākäsä* ‘set on fire’ : *tāk^wäsä* ‘shoot’; *käsä barä-* ‘have dysentery’ : (*a*)*k^wäsä* ‘remove the layers of the trunk of the *āsāt*-plant’.

k is in alternation with *q* of the other Gurage dialects: C. *korbeša* ‘the young of a goat’, Ed. *qorbešša*; C. *ākäsä* ‘be delayed’, So. *iqqäsä*; C. *k^wätä* ‘loft’, En. *q^wätä*; C. *kārä* ‘day’, M. *qänä* (for *n* : *r*; see § 20.7.3.1).

Original *k*, *k^w* has a tendency of becoming a spirant *x*, *x^w*. There are nearly as many nouns and verbs with *k* as there are with *x*.

Initial position: C. *xarä* ‘know’, root *khl*; C. *xənnä* ‘cubit’, So. *kərrä*; C. *xabä* ‘do something again’, G. *kä'abä*; C. *xäpt* ‘liver’ (§ 20.14.1.3), G. *käbəd*; C. *x^wä* ‘spill’, G. *kä'awä*; C. *x^wäx^wäb* ‘star’, G. *kokäb*.

In medial position: C. *äxər* ‘cereal’, root *äkəl*; *bäxər* ‘first-born’, root *bäkər*; *naxä* ‘send’, Amh. *lakä* (for *l* : *n*, see § 20.7.1.2); *tä-xätärä* ‘put on a dress’, root *kdn* ‘cover’; *fänäxä* ‘be patient’, So. *färräkä*.

In final position: C. *sox* ‘thorn’, So *äsok*.

Before or after a consonant: C. *mäsxäro*, name of the first month, Amh. *mäskäräm*; *šäxra* ‘clay’, So. *šäkla*.

In the verbal class *Cky*, *k* may become palatalized into *kʲ*: thus, C. *bäkʲä* ‘cry’, root *bky*; C. *säkʲä* ‘flee’, root *sky*.

k may become palatalized into *č*: thus, C. *wärčä* ‘front of leg’ (also Amh. *wärč*), from **wärk-e*; C. *čəff*, interjection used to chase away a cat, Amh. *kəff*.

For *k* becoming labialized *kʷ*, see § 20.2.

C. *kʷä* may become *ko* in the other Gurage dialects: thus, C. *kʷäfita* ‘plaited turban worn by Muslims’, E. *kofita*.

20.3.3. *x*

x most often comes from *k* (see § 20.3.2).

It also corresponds to *h* in Cushitic loanwords (as in C. *xäbər* ‘uncultivated pasture land’, from Qab. *haburra*), and to *h*, *ḥ*, and *ḥ* in Arabic loanwords (as in C. *xawa* ‘empty’, Ar. *hawā*; *saxən* ‘plate’, Ar. *ṣaḥn*; *xädämä* ‘render service’, Ar. *ḥadama*; *xamis* ‘Thursday’, Ar. *ḥamīs*).

Minimal pairs *x* : *xʲ* in C. *tä-xätärä* ‘put on a dress’ : *tä-xʲätärä* ‘follow’; *mäxäna* ‘six-month period of growing’ : *mäxʲäna* ‘bull in his prime’.

Minimal pairs *x* : *xʷ* in C. *xəm* ‘that’ : *xʷəm* ‘thousand’.

For *x* labialized into *xʷ*, see § 20.2.

For *x* palatalized into *xʲ*, see § 20.9.1.

In the verbal class *Cky*, the final *kʲ* may become *xʲ* in some nominal forms: e.g., C. *bixʲä* ‘mourning’, from *bäkʲä*.

C. *xʷä* may become *xo* and C. *xʷə* may become *xu* in the other Gurage dialects: thus, C. *xʷärinä* ‘dress made of hare’s skin’, Gt. *xorinä*; C. *xʷəm* ‘thousand’, Gt. *xum*.

20.3.4. *q*

q corresponds to *q* of Semitic and Ethiopic: e.g., C. *qomä* ‘stand’, Ar. *qāma*, G. *qomä*.

Minimal pairs *q* : *qʷ* in C. *qänärä* ‘be light’ : *qʷänärä* ‘finish the top of the roof’; *qəyä* ‘stone delimiting the boundary’ : *qʷəyä* ‘ball made of cloth or fiber’.

Minimal pairs of *q* : *q'* in C. *qäpärä* 'help' : *q'äpärä* 'plant'; C. *qätärä* 'kill' : *q'ätärä* 'join by attaching'; C. *bäqärä* 'germinate' : *bäq'ärä* 'brew beer'.

C. *q* alternates with *k* of the other Gurage dialects: thus, C. *qəṭär* 'leaf', En. *kä'är*; C. *qanča* 'fiber of the *äsät*-plant', Ed. *ke'ä*; C. *wäqt* 'time' (from Arabic), W. *wäkt*.

For *q* becoming a labialized *q''*, see § 20.2.

For *q* becoming palatalized *q'*, see § 20.9.1.

For Ethiopic *q* alternating with *k* of Chaha, see § 20.3.2.

q' of Chaha may become *ç* in the other dialects: e.g., C. *aq'ä* 'crunch', but So. *aççä*; C. *q'ätä* 'be weak', but W. *çetä*.

C. *q"ä*- as against *qo*-, and C. *q"ə*- as against *qu*- in the other Gurage dialects: thus, C. *q"ämärä* 'become strong', En. *qomärä*; C. *q"ərər* 'light', En. Ed. *qurər*.

C. *qə*- (probably from *q"ə*-) as against *w* in the other dialects; thus, C. *qənçuwät* 'woman who has power of stopping rain', Gt. *wīnçuwät*.

20.4. Dentals

The dentals are: *d*, *t*, *ṭ*.

20.4.1. *d*

d corresponds to *d* of Ethiopic. Examples: *däfärä* 'dare' (Amh. *däffärä*).

In Arabic loanwords *d* corresponds to *d*, *ḍ*. Examples: C. *xädämä* 'support one's parents', from Ar. *ḥadama* 'serve'; *qabd* 'object taken as a guarantee', from Ar. *qabḍ*.

Minimal pairs of *d* : *ḡ* in C. *däpärä* 'add, join' : *ḡäpärä* 'accomplish'.

By assimilation to a glottalized, *d* has become *ṭ*: thus, C. *wätäqä* 'fall' (Amh. *wäddäqä*); see § 20.14.1.5.

By assimilation to a voiceless *p*, *d* became *t* in C. *xäpt* 'liver', from *käbd* (see § 20.14.1.3).

For *t-g* of the various Gurage dialects becoming *d-g* in Chaha, as in *dəḡär* 'hair' (but Z. *ṭəḡär*), *däg'ä* 'honeyed water' (but W. *ṭəgay*), and for *t-d* becoming *d-d*, as in C. *mədad* 'griddle' (but in So. *məṭad*), see § 20.14.1.5.

For an originally geminated *d* becoming a devoiced *t*, see § 20.10.3.

20.4.2. *t*

t corresponds to *t* of Semitic and Ethiopic: C. *täfa* 'spit', Ar. *taffa*, Amh. *täffa*.

In Arabic loanwords Chaha *t* corresponds to *t*, *ṭ*: e.g., C. *tarik* 'story' (Ar. *ta'rīḥ*); C. *tumma* 'garlic' (Ar. *tūm*).

Minimal pairs of *t* : *č* in C. *tərātārā* ‘unroll thread used for weaving’ : *čərāčārā* ‘practice small trade’.

Original *t* of Chaha alternates with *ṭ* of the other Gurage dialects: thus, *tata* ‘twist two or more threads’, Go. So. *taṭṭa*.

Original *ṭ* of Ethiopic alternates with *t* of Chaha: C. *atebāt* ‘finger’ (Ms. *aṭebāt*; Amh. *ṭat*); see also § 20.11.2.

By assimilation to a palatal, *t* may become *č*: e.g., C. *gučəča* ‘earring’ (En. *gutəča*).

For the palatalization of *t* see § 20.9.1.

20.4.3. *ṭ*

ṭ corresponds to 1) Ar. *ṭ*, G. Te. Tna. *ṣ*, South Ethiopic *ṭ*: e.g., C. *ṭāma* ‘be thirsty’, Ar. *ṭami’a*, G. Te. *ṣām’a*, Amh. *ṭamma*.

2) Ar. *ḍ*, G. *ḍ*, Te. Tna. *ṣ*, South Ethiopic *ṭ*: e.g., C. *ṭāmādā* ‘yoke’, Ar. *ḍamada*, G. *ḍāmādā*, Te. *ṣāmda*; Amh. *ṭāmmādā*.

3) Ar. *ṭ*, North and South Ethiopic *ṭ*: C. *ṭāp’ā* ‘suck’, Ar. *ṭiby* ‘teat’, G. *ṭābāwā* ‘suck’, Am. *ṭābba*.

Minimal pairs of *ṭ* : *č* in *ṭəra* ‘grudge’ : *čəra* ‘fly whisk’.

The glottalized *ṭ* of Chaha alternates with the original *t* of the other languages: thus, C. *dəṣṭ* ‘kind of pot’ (So. *dəst*); *šāffāṭā* ‘become a rebel’ (Ed. *šāffāṭā*).

For the glottalized *ṭ* of C. *arāqəṭ* ‘leech’ in relation to *alāqəṭ* of the other Ethiopian languages, and of C. *wāṭāqā* ‘fall’ in relation to *wdq* of the other languages, see § 20.14.1.5.

For the palatalization of *ṭ* to *č*, see § 20.9.1.

20.5. Sibilants

The sibilants are: *z*, *s*, (*š*).

20.5.1. *z*

z corresponds to *z* of Semitic: e.g., C. *zābār* ‘time’, Ar. *zamān*, G. *zāmān*.

It also corresponds to Ar. *ḍ*, Hebr. *z*, Ethiopic *z*: e.g., C. *zəmb* ‘fly’, Ar. *ḍubāb*, Hebr. *zəḥūḇ*, Amh. *zəmb*.

In Arabic loanwords *z* corresponds to Ar. *z*, *ḍ*, *ṭ*: e.g., C. *ḡeza* ‘gratitude’, Ar. *ḡazā*; C. *kəzəb* ‘lie’, Ar. *kiḍḇ*; C. *zəxur* ‘midday’, Ar. *ṭuhr*.

Minimal pairs *z* : *ž* in *zānārā* ‘jump’ : *žānārā* ‘block the view with a curtain’.

For the palatalization of *z* into *ž*, see § 20.9.1.

An originally geminated *z* has become a devoiced *s*: e.g., C. *bäsa* ‘abound’, E. *bäzza*; *zasa* ‘act mad’ as against E. *zazza* (see § 20.10.8).

The voiced *z* remained *z* in the verbs with identical second and third radicals: thus, C. *fäzäzä* ‘be better’, E. *fäzzäzä*.

It also remained *z* in Amharic loanwords: e.g., C. *tä-razäzä* ‘make the last will’ from Amh. *tä-nazzäzä*.

For the palatalization of *z* to *ž*, see § 20.9.1.

20.5.2. *s*

s corresponds to 1) Ar. *ṣ*, Hebr. *š*, G. *ś*, other Ethiopian languages *s*: e.g., C. *sost* ‘three’, Ar. *ṭalāt*, Hebr. *šāloš*, G. *śālās*, Amh. *sost*.

2) Ar. *ṭ*, Hebr. *š*, Ethiopic *s*: e.g., C. *samt* ‘eight’, Ar. *ṭamān*, Hebr. *šəmonē*, Amh. *səmmənt*.

3) Ar. *s*, Hebr. *š*, Ethiopic *s*: e.g., C. *sämma* ‘hear’, Ar. *sami’a*, Hebr. *šāma*, G. *sām’a*.

4) Ar. *š*, G. *ś*, other Ethiopian languages *s*: e.g., C. *säpa* ‘be fatter than expected’, Ar. *šabḥ* ‘large’, G. *śabḥa* ‘be fat’, Amh. *säbba*.

In Arabic loanwords C. *s* corresponds to Ar. *s*, *ṣ*: e.g., C. *säbäb* ‘cause’, Ar. *sabab*; C. *asər* ‘time of day about 4 o’clock in the afternoon’, Ar. *‘aṣr*.

Minimal pairs *s* : *š* in C. *sägärä* ‘amble’ : *šägärä* ‘change’.

For *s* going back to a geminated *z*, see § 20.10.8.

For *s* palatalized into *š*, see § 20.9.1.

20.5.3. *š*

š is preserved in loanwords from Ge‘ez through Amharic: e.g., *šälot* ‘prayer’, Amh. *šälot* (from Ge‘ez).

Ge‘ez *š* has become *ʈ* in Chaha.

20.6. Laryngeal *h*

h, the only laryngeal in Chaha, exists in a few Amharic loanwords: *haymanot* ‘belief’, *har* ‘silk’.

20.7. Liquids

The liquids are: *l*, *n*, *r*.

20.7.1. *l*

There are a few lexemes in Chaha that have a simple or a geminated *l*. Nearly all the verbs and nouns that have retained *l* are loanwords from Amharic, Arabic, or Cushitic.

In initial position: *ləbəč barä* ‘become lukewarm’ (probably from Cushitic); *lul* ‘pearl’ (from Ar. *lu’lu*); *lämča* ‘twin’ (from Oromo).

Medial position: *agälgəl* ‘basket’ (from Amharic); *bučəlla* ‘puppy’ (from Oromo through Amharic); *atäla* and *atära* ‘lees of beer’ (from Amharic); *ella* ‘covet’ (from Cushitic).

Post-consonant position: *abläšä* (also *abräšä*) ‘spoil’ (from Amharic).

20.7.1.1. *Loss of original medial l*

Original medial *l* is occasionally lost through palatalization or through weakening of *l*. Examples: C. *gef* ‘tall’ (root *glf*); *beṭ* ‘clever’ (root *blṭ*); *eb* ‘milk’ (from **alib*); *gaba* ‘trodden straw’ (Go. *gäläba*).

Normally, original *l* became *n* or *r* (see below).

20.7.1.2. *l > n*

The original liquid *l* has become *n* in initial position and in medial position when originally geminated.

Initial position: *naba* ‘waist’ (M. *laba*); *näma* ‘flourish’ (Amh. *lämma*).

Medial position: *m^wäšəna* ‘kind of millet’ (Amh. *mašəlla*); *bäna* ‘eat’ (Amh. *bälla*); *säna* ‘arrive’ (M. *sälla*); *anäbä* ‘to milk’ (Amh. *alläbä*).

20.7.1.3. *l > r*

The liquid *l* becomes *r* in non-initial position when originally not geminated.

In intervocalic position: C. *ənqura* ‘egg’ (Ed. *ənqulä*); *gamera* ‘camel’ (Amh. *gəməl*); C. *qäpärä* ‘decrease’ (S. *qäbälä*).

Pre-consonant position: *darka* ‘dewlap’ (M. *daləgga*); *wärd* ‘child’ (root *wld*).

Post-consonant position: *amre*, name of the eleventh month (Amh. *hamle*); *a-mrana* ‘be slippery’ (M. *a-mlalla*); *a-čranämä* ‘be dark’ (M.a-člallämä).

Final, post-vocalic position: *bar* ‘holiday’ (Amh. *bal*); *ambər* ‘cabbage’ (S. *aməḷ*; for inserted *b*, see § 20.12.3).

20.7.1.4. *l > y*

When preceded by a palatalizing vowel, *l* is palatalized into *y*: thus, *gədiya* ‘sleep’ (n.), from *gätärä*, root *gdl*; *dəbiya* ‘addition’, from **dbl*.

20.7.2. *n*

n remains *n* in initial position: *näpärä* ‘it was’; *nəb* ‘bee’.

n remains *n* in intervocalic position when originally geminated: C. *anäsä* ‘be little’ (Amh. *annäsä*); *āfuna* ‘nose’ (E. *āfunna*).

A non-geminated *n* in intervocalic position remains *n* in Amharic and Cushitic loanwords: *fənāda* ‘explode’ (Amh. *fānādda*); *dina* ‘enemy’ (Oromo *dina*); *eyana* ‘luck’ (Kam. Had. *ayana*).

It remains *n* in medial, pre-consonant position: *andər* ‘small drum’; *čanqʷa* ‘not yet ripe’.

It likewise remains *n* in loanwords in final post-vocalic position: *saxən* ‘dish’ (from Ar. *ṣaḥn*).

A non-geminated *n* in intervocalic position becomes *r*: e.g., *amärä* ‘believe’ (Amh. *ammänä*); *qärä* ‘bladder’ (M. *qäna*); *arat* ‘top of head’ (M. *anat*).

In final postvocalic position *n* becomes *r*: e.g., *ənzər* ‘ear’ (M. *ənzən*, W. *əzən*); *gunär* ‘head’ (M. *gunnän*).

In verbs ending in *n* in the various Gurage dialects (often going back to a verb ending in *y*) Chaha has *n*: thus, *tänä* ‘swear’, but En. Gt. *tāñä*, Ed. *tāññä* (root *try*); C. *qʷänä* ‘roast grain or coffee’, but En. *qʷāñä*, Ed. *qʷāññä*; C. *a-fʷänä* ‘take a rest’, but En. Gt. *a-fʷāñä*, S. *a-fōye*; C. *tä-kranä* ‘rent’, but En. Gt. *tä-krañä*, Ed. *tä-krayyä*, Amh. *tä-kārayyä*.

The nasal *n* may be absorbed in the nasalized vowel: *äxʷärä-n* ‘immature’ (for *an-xʷärä-n*, lit. ‘it is not for him’); *āfuna* ‘nose’ (root *ñf*).

20.7.3. *r*

An original non-geminated *r* remains *r*.

In medial position: *däfärä* ‘dare’ (Amh. *däffärä*); *atärä* ‘spend the night’ (Amh. *addärä*; for *t:d*, see § 20.10.3).

In pre-consonant and post-consonant position: *bärča* ‘misfortune’; *boträka* ‘crack in the ground after the rainy season’.

In final position: *ägər* ‘foot’.

20.7.3.1. *r* > *n*

The liquid *r* becomes *n* in initial position and in non-initial position when originally geminated.

Initial position: *noṭä* ‘run’ (Amh. *roṭä*); *nesa* ‘corpse’ (Amh. *resa*).

Medial position, originally geminated: *tänäqä* ‘be dry’ (Amh. *därräqä*); *bändä* (for *bänädä*) ‘be cold’ (Amh. *bärrädä*); *bəna* ‘clear weather after rain’ (Ms. *bərra*).

In loanwords *r* either remains *r* or it becomes *n*: *rux* ‘spirit’ (but also *lux*, from Ar. *rūḥ*), *qāna* ‘read the Koran’ (from Ar. *qara’a*).

20.7.3.2. *r > y*

Palatalization of *r* into *y* occurs in the verb and in the noun. Examples: C. *qəyā* ‘wait’ (W. *qerā*, Har. *qēraḥa*); *səyā* ‘buy’ (Go. *səṛā*); *tā-wyanā* and *tā-wranā* ‘be confused in making a decision’; *a-šrānā* and *a-šrāyā* ‘distinguish’; *wanṭiyā* ‘bamboo filter’ (from *nāṭārā*); *ang^wāya* ‘not yet ripe (coffee, fruit)’, Ms. *ang^war*.

The liquid is also lost through palatalization in the imperfect 2nd feminine singular: *təgāfi* from *gāfārā* ‘release’; *təbādi* from *bātārā* ‘advance’.

There are occurrences where *r* is lost in medial position: *qem* ‘gleanings’ (W. *qārma*); *čəyā* ‘melted butter that runs down from the face of a woman’s hair’ (M. *čəṛəyā*).

20.7.3.3. *Other r processes*

By pseudo-correction (see § 20.17) *r* has become *l*: *sulle* ‘kind of breeches’ (Amh. *surre*); *fällämä* ‘sign’ (Amh. *färrämä*).

For the insertion of *r*, see § 20.12.2.

20.8. Palatals and palatalization

The palatals are : dentals and sibilants *ḡ ḥ ḥ̣ ḥ̣̣ ḥ̣̣̣*; velars *g’ k’ x’ q’*; and *y*. Note that palatal velars occur in West Gurage and in Muher, less in Masqan and Gogot. The number of the palatal velars in So., W., and Selti is small.

For originally geminated *ḡ* becoming *č*, for *ḥ̣̣̣* becoming *š*, and for *g’* becoming *k’*, see § 20.10.

20.8.1. *Minimal pairs*

d : *ḡ* in *dāpārā* ‘add, join’ : *ḡāpārā* ‘accomplish’.

t : *č* in *təṛātārā* ‘unroll thread used for weaving’ : *čəṛāčārā* ‘practice small trade’.

t : *č* in *təra* ‘grudge; *čəra* ‘fly whisk’.

s : *š* in *sāḡārā* ‘amble’ : *šāḡārā* ‘change’.

z : *ḥ̣̣̣* in *zānārā* ‘jump’ : *ḥ̣̣̣nārā* ‘block the view with a curtain’.

g : *g’* in *gāra* ‘buttocks’ : *g’āra* ‘leaf of the *āsāt*-plant that is torn’; *gātārā* ‘put to sleep’ : *g’ātārā* ‘be lenient toward someone’.

k : *k’* in *wāka* ‘beam’ : *wāk’a* ‘loan of a cow’.

x : *x'* in *tā-xātārā* 'put on a dress' : *tā-x'ātārā* 'follow'; *māxāna* 'the six-month period of growing and harvesting' : *māx'āna* 'bull in his prime'.

q : *q'* in *qāpārā* 'plant' : *q'āpārā* 'help'; *qāṭārā* 'kill' : *q'āṭārā* 'join by attaching'.

20.8.2. Alternation of palatals

č : *š*, *š* : *č* in C. *čəmbəra* 'chickpea' : M. *šəmbura*; *g'āčā* 'appointment' from Sid. *gēšša*; C. *wərš*, interjection used to chase away a calf, Ed. *wərč*.

ž : *ǵ* in C. *q'āžā* 'defecate' : Z. *qūǵi*.

20.9. Palatalization

It was seen above (§ 20.8) that the palatals are phonemes. They are also, however, phonetically conditioned. As a rule, a palatal consonant is brought about by the presence of a final *-i*, *-e*, and of a semivowel *y*. The final *-i* occurs in the imperfect, jussive, and imperative of the 2nd feminine singular, in the participle, and in the impersonal. The palatalizable consonants are the dentals *d*, *t*, *ṭ*, the sibilants *z*, *s*, and the velars *g*, *k*, *x*, and *q*.

20.9.1. Examples

d > *ǵ*: *tərāmǵ* 'you (fem.) love' (from *nāmādā*); participle *zamaǵ* 'runaway horse' (from *zāmādā*); impersonal *ag'āǵim* 'one binds' (from *agādā*).

t > *č*: *təkāfči* 'you (fem.) open' (from *kāfātā*); *āčoči* 'which one plowed' (from *čotā*).

ṭ > *č*: *təṭābči* 'you (fem.) seize' (from *ṭābāṭā*); *yarči* 'which one slaughters' (from *anṭā*).

z > *ž*: *tačānž* 'you (fem.) crouch down' (from *a-čānāzā*); impersonal *tā-g'ag'āžim* 'one is haughty' (from *tā-g'ag'āzā*).

s > *š*: *təqāmš* 'you (fem.) taste' (from *qāmāsā*); impersonal *nāk'āšim* 'one bit' (from *nākāsā*).

k > *k'*: *təsāk'ək* 'you (fem.) drive a peg into the ground' (from *sākākā*); note the palatalized first *k'* by assimilation.

x > *x'*: *təfār^x* 'you (fem.) are patient' (from *fānāxā*).

q > *q'*: *təṭār^q* 'you (fem.) draw water from a container' (from *ṭānāqā*). Nominal pattern: *wəṭaq'ā* 'act of falling down' (from *wāṭāqā*).

20.9.2. Verbs with 3rd radical *y/w*

In the verbs in which the 2nd radical is palatalizable and the 3rd radical is *y*, *w*, the 2nd radical becomes palatalized and *y*, *w* is absorbed in the

palatalization. Examples: *ačä* ‘close’ (G. *‘ašāwā*), *a-räšä* ‘build’ (G. *räsäyā* ‘place’), *ažä* ‘see’ (root *hzy*), *bäk’ä* ‘cry’ (G. *bākāyā*), *aq’ä* ‘crunch’ (root *hgy*).

The same feature occurs in the other Gurage dialects: *d* > *ğ* as in E. *ağğä* ‘sweep’ (G. *‘adāwā*); *t* > *č* as in S. *säče* ‘drink’ (G. *sätäyā*); *k* > *č* as in S. *bäče* ‘cry’ (G. *bākāyā*).

For the palatalization of the liquids, see § 20.7.1.4, § 20.7.3.2.

20.10. Devoicing

The voiced *b*, *bʷ*, *d*, *g*, *gʷ*, *ğ*, *z*, *ž* if originally geminated have normally become devoiced. Thus, *bb* > *p*, *bbʷ* > *pʷ*, *dd* > *t*, *gg* > *k*, *ggʷ* > *kʷ*, *ğğ* > *č*, *zz* > *s*, *žž* > *č*.

20.10.1. *b*

Originally geminated voiced *b* has become devoiced *p*. Verbs: *däpärä* ‘add’, E. *däbbärä*; *šäpätä* ‘choose, prefer’, E. *šäbbätä*; *gäpa* ‘enter’, E. *gäbba*.

Nouns: *ğipä* ‘mat made of fiber’, E. *ğibbä*; *dəpeya* ‘stick for hitting a ball’, E. *dəbbe*.

In final, postvocalic position: *top*, interjection used to chase away a horse, E. *tobb*; *žäp* ‘lion’, E. *žäbb*.

Note also the devoicing of *b* in *xäpt* ‘liver’, from *käb(ə)d*, by assimilation to the final *t* (see also § 20.14.1.3).

The originally geminated *b* remains *b* in Amharic loanwords: e.g., *a-näbäbä* ‘read’ from Amh. *a-näbbäbä*; *g’äbärä* ‘pay taxes’ from Amh. *gäbbärä*, and others.

20.10.2. *bʷ*

Originally geminated voiced *bʷ* has become devoiced *pʷ* in Chaha *kapʷat* ‘mattress made of unsplit rope and fiber’ as against E. *kabbʷat*.

20.10.3. *d*

Originally geminated voiced *d* has become devoiced *t* in Chaha. Verbs: *gätärä* ‘put to sleep’, E. *gäddärä*; *bätärä* ‘advance’, E. *bäddärä*.

Nouns: *agdata* ‘partition in the house’, E. *agdadda*; *data* ‘chest’, E. *dadda*.

In verbs with identical second and third radicals the original *d* remains *d* in order to avoid the succession *t-d*: thus, *sädädä* ‘send’ (however En. Gt. have *sätädä*, Ed. *sättädä*); *gädädä* ‘tear’ (but En. Gt. *gätädä*, Ed. *gättädä*).

It so happens that in verbs with final *r*, Chaha has preserved the voiced *d*: thus, *xädärä* ‘thatch a house’ (but En. Gt. *xätärä*); *nädärä* ‘bore a hole’ (but Ed. Gt. *nätärä*).

Chaha has also kept the original *d* in Amharic loanwords: *näda* ‘help’ (Amh. *rädda*; for *r* : *n*, see § 20.7.3.1); *tä-ṭadäfä* ‘hurry’ (Amh. *tä-ṭaddäfä*).

There are also doublets: *adärä* and *atärä* ‘spend the night’; *adägä* and *atägä* ‘throw away’.

20.10.4. *g*

Originally geminated voiced *g* has become devoiced *k*. Verbs: *gäkämä* ‘hit with the fist or elbow’, E. *gäggämä*; *mäkarä* ‘suppurate’, E. *mäggärä*; *näka* ‘coagulate’, E. *nägga*, Amh. *rägga*.

Nouns: *däk* ‘calf of a certain age’, E. *dägg*; *gäkäbät* ‘chest of animal’, E. *gäggäbät*.

The presence of another voiced in the root prevented the devoicing of *g* into *k*: thus, *zägädä* ‘remember’, but En. Gt. *zäkädä*; *agädä* ‘bind’, but En. Gt. *akädä*.

The voiced *g* is also preserved in Amharic loanwords: *agäzä* ‘help’, Amh. *aggäzä*; *sägädä* ‘worship’, Amh. *säggädä*; *nägäsä* ‘reign’, Amh. *näggäsä*.

20.10.5. *g'*

Originally geminated voiced *g'* has become devoiced *k'*. Verbs: *g'äk'ätä* ‘accompany’, E. *g'ägg'ätä*; *g'äk'ärä* ‘straighten out’, E. *g'ägg'ärä*.

Nouns: *wək'är* ‘evening gathering’, E. *wəgg'är*; *muk'ik'a* ‘grass growing in the *äsät* plantation in the rainy season’, E. *mug'igg'a*.

20.10.6. *g^w*

Originally geminated voiced *g^w* has become devoiced into *k^w*. Examples: *gäk^wärä* ‘become flexible having been put on the fire’, E. *gägg^wärä*; *gäk^wä* ‘main root of the *äsät*-plant’, M. *gogg^wä*; *gäk^wärä*, expression of thanks to God having eaten something for the first time, E. *gäk^wärä*.

20.10.7. *ǧ*

Originally geminated voiced *ǧ* has become devoiced *č*. Examples: *gačä* ‘rope an animal to kill it’, E. *gaččä*; *məğäča* ‘fireplace’, E. *məğäččä*.

20.10.8. *z, ž*

Originally geminated voiced *z* becomes *s*, and geminated *ž* becomes *š*. Examples: *bäsa* ‘abound’, E. *bäzza*; *zasa* ‘act mad’, W. *zazza*; *zəməsämä* ‘be

wet', E. *zəməzzämä*; *zəsər* 'span', E. *zəzzər*.

The voiced *z* has remained *z* in verbs with identical second and third radicals: thus, *fäzäzä* 'be better', E. *fäzzäzä*; *bazäzä*- 'be in low spirits', E. *bazzäzä*-.

It has also remained *z* in Amharic loanwords: *tä-razäzä* 'make the last will', from Amh. *tä-nazzäzä* (for *n : r*; see § 20.7.2); *näzäbä* 'be flexible', from Amh. *läzzäbä* (for *l : n*; see § 20.7.1.2).

Examples for **žž* becoming *š*: *žäšä* 'be cold', E. *žäžžä*; *gašä* 'raid', E. *gažžä*; *ašä* 'see', E. *ažžä*; *bräšä* 'bolt and run away', E. *bäräžžä*.

20.11. Glottalized and non-glottalized

There are several examples of alternation between *q:k*, *t:t*, and *č:č* regardless of whether the original consonant is the glottalized (*q*, *t*, *č*) or the non-glottalized (*k*, *t*, *č*). While in some examples the alternation of glottalized versus non-glottalized could be explained through assimilation as may be the case of *atebät* 'finger' (M. Ms. Go) as against *atebät* (C. E), that is, *t-t* becoming *t-t*, this is not the case of the other examples. For a possible Cushitic influence, see Leslau 1952.

20.11.1. *k : q*

Original *k* in alternation with *q*: C. *äkäsä* 'be delayed', So. *iqqäsä*; C. *korbeša* 'the young of a goat', Ed. *qorbeššä*.

Original *q* in alternation with *k*: E. *qurä* 'crow', C. *kurä*; C. *qanča* 'fiber of the *äsät*-plant', Ed. *ke"ä* (for the alternation of glottalized with the glottal stop, see Leslau 1992: 82–84).

20.11.2. *t : t*

Original *t* alternates with *t*: thus, C. *šäffätä* 'become a rebel', but Ed. *šäffätä*; M. *aläqet* 'leech', but C. *aräqet* 'leech' (possibly through assimilation of glottalization, namely *q-t* becoming *q-t*); C. *atebät* 'finger', but So. *aṭabät*.

20.11.3. *č : č*

In C. *čäfa* 'marshland', but En. *čäfa*; C. *čära* 'liquid mud', but En. *čära*; C. *čəyä* 'hump of a cow', but En. *čəyä*; C. *a-x'ačä* 'encourage', but Ed. *a-xeččä*.

20.12. Insertion of a non-etymological *n*, *r*, *b*

Insertion of these consonants occurs mostly in roots with a velar *g*, *k*, *q*. The majority of occurrences with inserted *n*, *r* occurs in East Gurage.

20.12.1. *n*

Examples of Chaha roots with inserted *n*: *grangər* ‘the young of an animal’ (Amh. So. *gəlgəl*); *atankərt* ‘eucalyptus’ (So. Amh. *atakəlt*); *mānkina* ‘car’ (Amh. *mākina*); *xendä* ‘tread’ (G. *kedä*); *ənk’ak’ä* ‘scab’ (C. *akākä* ‘scratch when it itches’); *ənzər* ‘ear’ (W. *əzən*); *genzo* ‘ax’ (Ms. *gezo*); *g”änčä* ‘hyena’ (E. *g”äčä*); *sočä* ‘part of the tapeworm that emerges after medication’ (E. *sočä*); *wägängä* ‘bamboo spatula’ (M. *wälgäḡa* > **wägäḡa*); C. *wanqäba* ‘wooden plate for measuring coffee’, E. M. *waq’äma*.

There are occurrences of the prefix *ən-*: *ənqura* ‘egg’ (Amh. *qura* ‘testicle’); *ənqurquna* ‘excrement of a sheep or goat’ (So. *q”äräq”ärya*); *ənbərbäya* ‘butterfly’ (Amh. *birabiro*).

20.12.2. *r*

ənqərčä ‘a kick’ (from *näqäṭä* ‘kick’). Insertion of *n* and *r* occurs in *ənqurfit* ‘obstacle’ (G. *‘əqf-ät*).

20.12.3. *b*

The labial *b* after *m* is inserted only in *ambər* ‘cabbage’ (W. *aməl*); *qambisa* and *qāmis* ‘shirt’ (*qāmis* is borrowed from Arabic).

20.13. Gemination

Chaha is a non-geminating dialect. Thus, the perfect has a non-geminated consonant (e.g., *nākäsä*) as against the other South Ethiopian languages in which the 2nd radical is geminated (e.g., *näkkäsä*). From a diachronic point of view, however, there are indications of an early stage of gemination. Indeed, whenever an originally voiced consonant is expected to be geminated it became voiceless. This is the case of *b*, *b”*, *d*, *g*, *g’*, *g”*, *ḡ*, *z*, *ž* having become voiceless *p*, *p”*, *t*, *k*, *k’*, *k”*, *č*, *s*, *š* (see § 20.10).

Occasionally one encounters occurrences with a geminated radical, as in *ella* ‘cove’, *at-ənnä* ‘put to sleep’, *tobbätä* ‘become Muslim’, *əmmat* ‘an only one’, *əkkəm* ‘in vain’, *ənnəm* ‘all’, *attəm* ‘any’. Some of these lexemes are loanwords.

Note also that gemination resulting from assimilation occurs in morpheme boundaries: e.g., *fəqunnät* ‘fatness’, from *fəqur-nät*; *bučənnät* ‘difference’, from *bučər-nät*; *fʷänniyä* ‘upward’, from *fʷär-niyä*; *yəxʷəllä* ‘it is possible for him’, from *yəxʷər-lä*; *bibənnä* ‘when he says to her’, from *bibər-na*; *säpännä* ‘we broke’, for *säpär-nä*.

Gemination also occurs when two homologous phonemes meet: *nəqqar* ‘much’ for *nəq qar*.

20.14. Assimilation

The consonants that are most susceptible to assimilation are the liquids, the labials, and the dentals. The glottalized consonants in the environment of non-glottalized consonants are either assimilatory or assimilated consonants.

In the partial assimilation voicing and unvoicing play a considerable role.

20.14.1. Partial assimilation

20.14.1.1. Assimilation of the dental *n* in a contiguous position

The dental *n* is assimilated and becomes a labial *m* when in contact with the labials *b*, *f*. In this combination *nb* > *mb*, and *nf* > *mf*.

nb > *mb*. Examples: Amh. *ənba* ‘tear’ > C. *əmba*; Amh. *gənb* (also *gəmb*) ‘stone wall’ > C. *gəmb*; Z. *gänbälä* ‘black’ > C. *gämbänä*.

nf > *mf*. Examples: C. *ənfačä* and *əmfačä* ‘nasal mucus’; **ənfas* ‘air, wind’ > C. *əmfas*.

20.14.1.2. Assimilation of the labial *m* in a contiguous position

The labial *m* is assimilated and becomes dental *n* in contact with the dentals *d*, *t*, *ʈ*, the sibilant *z*, the palatals *ʃ*, *č*, and the velar *q*. Thus, *md* > *nd*, *mt* > *nt*, *mʈ* > *nʈ*, *mz* > *nz*, *mš* > *nš*, *mč* > *nč*, *mq* > *nq*.

md > *nd*. Examples: E. *dumd* ‘joined’ > C. *dund*; E. *ʈəmd* ‘pair of oxen’ > C. *ʈənd*; E. *nəməd* ‘expensive’ > C. *nund*.

mt > *nt*. There are no examples of Chaha *mt* > *nt*, but this partial assimilation occurs in the other West Gurage dialects in relation to Chaha. Thus, C. *xumtət* ‘stiff, horny’ > Ed. *huntut*; C. *qʷämät* ‘gourd’ > Gt. *qōnt*.

mʈ > *nʈ*. E. *amʈ* ‘kind of grass’ > C. *anʈ*; En. *samʈ* ‘branch without leaves’ > C. *sonʈ*; Gt. *amʈarəyä* ‘dough that has a sour taste’ > C. *onʈarəyä*.

mč > *nč*. Example: En. *lumča* ‘kind of bush’ > C. *nunčä* (for *l* : *n*, see § 20.7.1.2).

mz > *nz*. Example: En. *gamziyä* ‘three fireplace stones on which the cooking pot rests’ > C. *gʷänziyä*.

mš > *nš*. This assimilation is attested in dialects other than Chaha. Thus, Z. *qolomšoši* ‘basket used for grain’, W. *q^wälänše*.

mq > *nq*. Example: Ed. *šəmqəq* ‘knot made in a belt’ > C. *šənnəq*.

20.14.1.3. Assimilation of voiceless-voiced > voiceless-voiceless

**xäbd* ‘liver’ > **xäpd* > *xäpt*.

20.14.1.4. Non-contiguous regressive assimilation of palatalization

A non-palatal followed by a palatal in a non-contiguous position becomes palatal. Examples: En. *gutəča* ‘earring’ > C. *gučəča*; En. *futfuča* ‘bread crumbs’ > C. *fučfučä*.

In the dialects other than Chaha: S. *liṭačča* ‘bark of tree’ > W. *ləččače*; E. *dəngər* ‘fat’, M. *gǎngən*; S. *mədağğa* ‘fireplace’, E. *məğəğğa*.

20.14.1.5. Assimilation of lexemes in which a glottalized is involved

t-g > *d-g* (glottal-voiced > non-glottal voiced) in Z. *ṭəgār* ‘hair’ > C. *dəgār*; W. *ṭəgay* ‘honeyed water’ > C. *däg’ä*.

q-d > *g-d* in G. *‘aqädä* ‘tie’ > C. *agädä*.

q-t > *q-ṭ* in M. *aläqət* ‘leech’ > C. *aräqət* (for *l*: *ṛ*, see § 20.7.1.3).

t-d > *d-d* in So. *məṭad* ‘griddle’ > C. *mədad*.

d-q > *ṭ-q* in S. *wädäqä* ‘fall’ > C. *wäṭäqä*; Ed. *däqäqä* ‘be ground finely’ > C. *ṭäqäqä*.

20.14.2. Total assimilation

Total assimilation occurs only when the two consonants are in contact. It is mostly regressive. The consonants that are frequently assimilated are the liquids *r*, *n*, but also other consonants.

20.14.2.1. *r*

r is assimilated to various consonants, mostly in dialects other than Chaha. Examples: C. *wärčä* ‘front leg of cattle’, Ed. *wäččä*; C. *qārčət* ‘basket made of bamboo’, En. Ed. *qäčät*; C. *gurda* ‘strong oath that should not be broken’, M. *gudda*.

rn > *nn* in morpheme boundaries: Gt. *fuqur-nät* ‘fatness’, C. *fəqunnät*; *bučər-nät* ‘difference’, C. *bučənnät*; **f’är-niyä* ‘upward’, C. *f’änniyä*.

20.14.2.2. *n*

n is assimilated to various consonants, mostly in dialects other than Chaha. Note that gemination that should have resulted from the process of

assimilation does not always occur. Examples: E. *kənčəf* ‘chin’, Ed. *kučəf*; C. *kānfār* ‘lip’, Ed. *kāfār*; C. *wānfīt* ‘sieve’, Z. *wāfīt*; C. *tənkiyā* ‘kind of jar’, Ed. *təkkiyā*; C. *maṇṭa* ‘liquid of the intestine’, Ed. *maṭā*.

20.14.2.3. *Liquids and labials*

In dialects other than Chaha: **angāt lā-dār* ‘upward slope’, Z. *angällädär*; E. *atšaḳətyä* ‘kind of *äsät*-plant’, En. *ašaḳət*; C. *fərfər* ‘kind of worm’, Ed. *fəffər*; M. *əmbuq balä* ‘take a mouthful of liquid’, So. *əbbuq balä* (unless M. *əmbuq* is for *ənbuq*, with inserted *n*, for which see § 20.12); Gt. *amṭarəyā* ‘dough that has a sour taste’, M. *oṭarəyā*; C. *käbsasa* ‘fiber of the *äsät*-plant’, En. *kasasa*.

20.15. Dissimilation

Dissimilation can be contiguous and non-contiguous, progressive and regressive. Dissimilation may also result in deglottalization and in voicing of one of the consonants.

20.15.1. *Progressive, non-contiguous*

t-q > d-q in Amh. *saqä* ‘laugh’, Gaf. *šaḳä* > **taḳä*, C. *daḳä*.

In dialects other than Chaha: M. *ṭarraḳqa* ‘moon’, So. *därraḳqa*; M. *aṭq* ‘joint of bamboo’, En. *andəq* (for inserted *n*, see § 20.12).

Dissimilation of voice resulting in voiced-voiced > voiced-voiceless. Examples: C. *tä-g’ag’äzä* ‘be proud’, but E. *tä-g’agg’äsä*.

Voiceless-voiceless > voiceless-voiced in M. *näqq’äsä* ‘limp’, but E. *näqq’äzä*; E. *a-čeqä* ‘make sure’ > **a-čekä* > *a-čegä*.

In verbs with medial *b* for which one would expect *p* in Chaha (see § 20.10.1), the voiced *b* is retained when preceded by a voiceless emphatic. Thus, C. *ṭäbäqä* ‘be tight’ instead of the expected *ṭäpäqä*; C. *ṭäbäsä* ‘roast’ instead of the expected *ṭäpäsä*.

20.15.2. *Progressive contiguous*

Dissimilation of voice resulting in voiced-voiced > voiced-voiceless in C. *dəm* ‘d’ ‘meeting place of two rivers’, but En. *dəm* ‘t’; C. *gung’ərəyā* ‘kind of knife’, but Gt. *gunčərəyā*.

20.15.3. *Progressive dissimilation*

Progressive dissimilation of two homorganic sounds either in Chaha in relation to the other Gurage dialects or in the various dialects outside of Chaha.

f-f > *b-f* as in E. *kəfäkkäfä* ‘thatch the roof with the tip of grass pointing up’, but En. *kəbäkäfä* (note Ed. *kəwäkkäfä*); C. *k’əfäk’äfa* ‘spray water’, but En. *k’əbäk’äfä*.

q-q > *b* (*w*)-*q*: C. *čəqäčäqä* ‘annoy by talking’, Gt. *čəbäčäqä*, Ed. *čəwäččäqä*.

g-g > *b-g* as in Amh. *dägäddägä* ‘level off the floor of the house’, En. *dəbätägä*.

x-x > *b-x*: C. *səxäsäxä* ‘shell corn by grinding slightly with a pestle’, En. *səbäsäxä*.

In the instrumental that is formed with the prefix *m*, the labial *m* became *w* in roots containing a labial. Thus, *wäfčä* ‘grinding stone’, *wäfteya* ‘spindle’ (root *fil*), *wägäpäčä* ‘horse-racing field’ (root *g’äpätä* ‘gallop, race’).

Note the same feature in some nouns without a labial: *wädrägya* ‘hammer’ (root *dng* ‘hit’ going back to *drg*); *wančiyä* ‘filter’ (root *ntr*).

20.15.4. Dissimilation of gemination

A geminated consonant may be dissimilated through the liquid *r*, occasionally through *n*. Examples: Ms. *mäddäd* ‘sickle’, but C. *märdäd*; E. *gədd* ‘misery’, but C. *gərd*; E. *k”ädda* ‘water bottle’, C. *k”ärda*; *šəšəra* ‘liquid that exudes from the *äsät*-plant’, also *šəršəra* (undoubtedly from **šəššəra*).

In dialects other than Chaha: W. *iddä* ‘carding bow’, En. *irda*.

20.16. Metathesis

Metathesis can be contiguous and non-contiguous. There are not enough examples in the same dialect, and, for that matter, not enough examples by comparing various dialects to warrant a definite statement on the phonetic conditions in which metathesis occurs.

20.16.1. Contiguous metathesis

The examples that follow are taken from Chaha as well as from the other Gurage dialects. Only a selection is given from the dialects other than Chaha.

Velar as 1st consonant: C. E. *əgzər* ‘God’, but Ms. Go. *əzger*; C. *əndäxrečä* ‘kidney’, but Ms. *əndärheččä*; C. *azgart* ‘cloth belt’, M. *azgaləd*, but Ms. *azlagəd*.

Liquid as 1st consonant: En. *bərčəqo* ‘drinking glass’, but C. *bəčərqo*; C. *arwa* ‘soul’, but W. *awrä*.

Dental as 1st consonant: C. *gʷādrāt* ‘mound around the fireplace’, but M. *gʷārdāt*; C. *aṭmet* ‘solidified juice of the white part of the *āsāt*-plant’, but En. *amīʾid* (for *ʾ* for *t*, see Leslau 1992: 84).

20.16.2. Non-contiguous metathesis

Velar as 1st consonant: So. *kāsāl* ‘charcoal’, C. *kāsār* (for *l*: *r*, see § 20.7.1.3), but Z. *sāhal*.

Labial as 1st consonant: C. *kābāro* ‘drum’, but Ms. *kārābo*; C. *wāfānča* ‘doorway’, but Ed. *wānāfč-ar* ‘section of the house from the entrance to the pillar’.

Liquid as 1st consonant: W. *gālāmā* ‘sew leather on the cover of the basket’, but C. *gāmārā* (for *l*: *r*, see § 20.7.1.3); Ms. *borkätta* ‘crack of the ground after the rainy season’, but C. *botrāka*.

Sibilant as 1st consonant: C. *kəzəb* (from Arabic) ‘lie’, but Ed. *kəbz*.

Palatal as 1st consonant: *fīšara* ‘sediment of water’, but Go. *fərša*.

20.17. Pseudo-corrections

There are a few occurrences of pseudo-corrections in the various dialects.

Since *b* may become *w* (see § 20.1.1) there are examples of *w* becoming *b*: thus, E. *kaba* ‘light, useless’, C. *kaba*, the correct form being C. *xawa*, from Ar. *ḥawā* ‘empty’; C. *bʷər* ‘main’ instead of *wər*.

The labial *m* may become *w* (see § 20.1.4). An example of *w* becoming *m* by pseudo-correction is M. Go. *tamuyä* ‘orphan’, to be identified with C. E. *tawəyā*.

For *q*’ becoming *č* (see § 20.3.4), one also finds examples of *č* becoming *q*’, as in C. *wanq’ābā* ‘wooden plate for measuring coffee’ going back to Ed. *wāčāba*, from Kamb. *wāčāba*; C. *q’āf barā* ‘be near to overflowing’ to be compared with Ed. *čəff barā*, Amh. *čāfāff alā*; C. *māq*’, interjection used to chase away a horse, but Ed. M. Amh. *māčč*.

The original *l* may become *r* in Chaha (see § 20.7.1.3). Occurrences of pseudo-corrections of *r* becoming *l* are: C. *fällāmā* ‘sign’ from an original *färrāmā*; C. *lux* ‘spirit’, from Ar. *rūḥ*.

20.18. Vowels

The vowels that occur in all the dialects are seen in Table 20-1.

Table 20-1. Chaha Vowels

	front	central	back
high	i	ə	u
mid	e	ä	o
low		a	

20.18.1. Additional vowels

Chaha has additional vowels: *ä* (open *ä*), *o* (open *o*); nasalized allophones *ã*, *õ*; allophone *â* of *ä*, and allophone *ũ* of *ə* when preceded by a labial.

Minimal pairs *ä* : *ẽ* in *bätärä* ‘be first’ : *bätärä* ‘distinguish, separate’; *wärät* ‘special share of a slaughtered cow’ : *wärät* ‘sleep’.

20.18.2. ä

Initial *ä* alternates with *ə* of Amharic: e.g., C. *äbät* ‘furuncle’, Amh. *əbät*; *äčä* ‘wood’, Amh. *ənčät*; C. *äxər* ‘cereal’, Amh. *əhəl*.

ä is a variant of *e* within Chaha and of *e*, *ə*, *i* of the other dialects. Examples: *wärät* and *wäret* ‘sleep’; *ärä* and *äre* ‘cows’; *etäniyā* and *eteniyā* ‘in which direction?’; C. *särəfər* and E. *särefər* ‘kind of breeches’; C. *däyā* ‘palm’, but E. *dəyā*, W. *diyā*.

ẽ is also contracted from *-ay*, as in C. *ozä* ‘evil eye’, but En. Gt. *ozay*; C. *sənä* ‘wheat’, but Gt. *sənay*.

ẽ also results from a combination *a-e*: thus, *yagdatä* ‘section of the house near the partition’ from *yä-agdata-e*.

20.18.3. o

Minimal pairs of *o* : *o* in *odä* ‘tell’: *odä*- ‘be necessary’.

o may result from the reduction of final *-äm*: thus, C. *mäsxäro*, name of the 1st month, Amh. Z. *mäskäräm*.

Initial *wä*, *aw*, *äw* may become *o*. Examples: C. *wägät* and *ogät* ‘assembly’; C. *od* ‘threshing field’ as against En. Gt. *awd*; C. *tona* ‘strength’, Gt. *täwna*.

o also results from the reduction of *aw-*, *awä-*: e.g., C. *oṭät* ‘one who leads the singing’, En. Gt. *awṭayt*; C. *ondä* ‘lower’, from *awändä*.

20.18.4. e

e also results from contraction of *ay*: e.g., C. *en* ‘eye’, Gt. *ayn*; *çet* ‘sun’, Gt. *çayt*; C. *eb* ‘milk’, S. W. *ayb*; C. *xäde* ‘roof thatcher’, E. *xäday*.

20.18.5. *i*

i also results from contraction of *-əyə*, *-äyə*, *-äəy*: e.g., *tiqäms* ‘while he tastes’ from *təyəqäms*; *tina* ‘with us’, from *tä-yəna*; *tiya* ‘with me’, from *ta-əya*.

20.18.6. *ə*

ə alternates with *i* of the other dialects: C. *təbre* ‘animosity between two people’, Ed. *tibre*.

20.18.7. *u*

u may go back to *əb* (passing through *əḥ*, *əw*): e.g., C. *tu* ‘breast’, Ms. *təb*, E. Ed. *təw*; C. *sur* ‘broken’, Go. So. *səbur*, C. also *səwə*.

20.18.8. Long vowels

Occasional long vowels occur in onomatopoeic expressions, such as *bā̄ barā̄*, *mā̄ barā̄* ‘bleat’; also in *ō* ‘yes’, *yō* ‘here I am’.

20.19. Vocalic initial

Vocalic initial goes back to Semitic and Ethiopic ’ ‘*h ḥ ḥ*. Examples: *äxər* ‘cereal’ (G. *ʾəxəl*); *äf* ‘bird’ (G. Te. *ʾof*); *äčä* ‘tree’ (G. *ʾəd*); *anä* ‘there is’ (G. *hallo*); *akäkä* ‘scratch’ (G. *ḥakäkä*); *adägä* ‘throw away’ (G. *ḥadägä*); *aməst* ‘five’ (root *ḥms*).

In Arabic loanwords: *axera* ‘future life’ (from Ar. *ʾahira*); *əmar* ‘donkey’ (from Ar. *ḥimār*); *amd* ‘thanks’ (from Ar. *ḥamd*).

In Cushitic loanwords: *äčba* ‘central pillar of the house’ (from Kam. *utubu*); *äfrax* ‘pregnant’ (from Had. *hafrakko*).

The initial vowel may be preceded by *y*. Thus, *yäčba* and *äčba* ‘central pillar of the house’; *yärbat* and *ärbat* ‘evening meal’; *yəra* ‘wet’, but Ms. *ira*; *yədar*, name of the third month, but Ms. W. *ədar*, *yəna* ‘we’, but En. Gt. *ina*.

20.20. Prothetic *ä*, *ə*

Prothetic *ä*, also *ə*, occurs before *r* and occasionally before *s*. Examples: *äram* ‘cow’, *äräq* ‘far’, *ärč* ‘boy’, *äro* ‘Wednesday’ (Amh. *rob*); *ərku* ‘impure’.

Before *s*: *äsər* ‘root’, *əset* ‘young grain’.

20.21. Meeting of vowels

Chaha avoids the coming together of two vowels. If, for morphological reasons, two vowels should come together, the following may occur: elision of one of the vowels, or introduction of a glide semivowel *y* if the preceding vowel is *u*.

20.21.1. Elision

a-e > *ä*: *yä-agdata-e* > *yagdatä* 'of the section of the house near the partition'

ä-ä > *ä*: *yä-äčba* > *yäčba* 'of the pillar'

ä-a > *a*: *yä-adot* > *yadot* 'of the mother'

ä-e > *e*: *yä-eb* > *yeb* 'of the milk'

ä-ə > *ə*: *yä-əmfas* > *yəmfas* 'of the wind'

20.21.2. Insertion of a semivowel

i-u > *i^yu*: *wami-u* > *wami^yu* 'it is (necessary) to do'

e-y > *e^yu*: *f^wäre-u* > *f^wäre^yu* 'it is above'

ä-u > *ä^yu*: *wäsä-u* > *wäsä^yu* 'it is necessary to buy'.

20.22. Syllabic structure

A word may begin with any vowel except with *i*: thus, *ab* 'father', *äčä* 'tree', *äšam*, form of greeting to welcome someone returning from a trip, *ebär* 'without', *əfəf barä* 'blow', *odä* 'tell', *əna* 'empty', *ux barä* 'sigh'.

A word may end with any vowel except with *u*:⁴ thus, *angača* 'cat', *ənkurkuyä* 'lumpy', *asä* 'coward', *ete* 'where?', *xə* 'that', *atawari* 'one who spies out', *angädo* 'cattle pen outside the house', *čuču* 'chick'.

Needless to say, some of the vowels either at beginning of the word or at the end of the word are rare.

For the reduction of syllables, see the individual vowels mentioned above.

20.23. Consonant cluster

Chaha has an initial consonant cluster only if the 2nd radical is *r*, the 1st radical being either the labial *b*, *f*, the velar *g*, *k*, rarely *d*. Examples: *bre* 'insect that eats human skin or cereal', *brät* 'iron, rifle'; *fräta* 'sprinkle water', *fräntä* 'break bread into pieces'; *grängər* 'the young of an animal', *greğät* 'young women of the same clan'; *krəstiyän* 'Christian', *kräta* 'lift up a heavy object'.

4. It is quite possible that initial *i* and final *u* occur in the language, but I did not record lexemes with these vowels in Leslau 1969.

Bibliography

- Cohen, Marcel. 1931. *Etudes d'éthiopien méridional*. Paris: Geuthner. [Pp. 104–33 deals with the phonology of various Gurage dialects.]
- Hetzron, Robert. 1972. *Ethiopian Semitic: Studies in Classification* (Journal of Semitic Studies Monograph 2). Manchester: The University Press.
- . 1977. *The Gunnän-Gurage Languages* (Istituto Orientale di Napoli. Ricerche 12).
- Leslau, Wolf. 1948. “Le problème de la gémiation du verbe tchaha (gouragué).” *Word* 4: 42–47.
- . 1950a. *Ethiopic documents: Gurage* (Viking Fund Publications in Anthropology 14). New York: The Viking Fund.
- . 1950b. “La position du gafat parmi les langues sémitiques de l’Ethiopie.” *Comptes rendus du groupe linguistique d’études chamito-sémitiques* 5: 47–48.
- . 1952. “The Influence of Sidamo on the Ethiopian Languages of Gurage.” *Language* 28: 63–81.
- . 1967. “The Impersonal in Chaha.” In *To Honor Roman Jakobson* (Janua Linguarum Series Maior 32), vol. 2, pp. 1150–62. The Hague: Mouton.
- . 1969. “Toward a Classification of the Gurage Dialects.” *Journal of Semitic Studies* 14: 96–109.
- . 1979. *Etymological Dictionary of Gurage (Ethiopic)*, vol. 3: *Etymological Section*, pp. xv–xcii. Wiesbaden: Harrassowitz. [See also 1992.]
- . 1983. *Ethiopians Speak: Studies in Cultural Background*, part 5: *Chaha-Ennemor* (Aethiopistische Forschungen 16).
- . 1992. *Gurage Studies: Collected Writings*. Wiesbaden: Harrassowitz. [Pp. 1–113, reproduction of 1979; pp. 117–20, reproduction of 1950a; pp. 138–40, reproduction of 1983; 8–10; pp. 430–42, reproduction of 1967.]
- Polotsky, Hans Jakob. 1938. “Etudes de grammaire gouragué.” *Bulletin de la société de linguistique de Paris* 38: 137–75.
- . 1951. Notes on Gurage Grammar (Oriental Notes and Studies 2). Jerusalem: Israel Oriental Society.
- . 1971. *Collected Papers*. Jerusalem: Magnes Press. [Pp. 477–515, reproduction of 1938; pp. 521–73, reproduction of 1951.]
- Shack, W. A. 1966. *The Gurage: A People of the Ensete Culture*. London: Oxford University Press.

Amharic Phonology

Wolf Leslau

University of California, Los Angeles

Amharic is the national language of Ethiopia. It belongs to the branch of the Semitic language family of which the other languages are Ge'ez (at present the language of the liturgy), Tigre and Tigrinya in the north (closely related to Ge'ez). The languages of the center and of the south are Amharic, Argobba (closely related to Amharic), Harari (or Adare, spoken only in the city of Harar), Gurage (a cluster of at least twelve dialects), and Gafat (for which the present writer found only four speakers in 1946).

In relation to the Ge'ez phonological system the Amharic phonology is characterized by the loss of the laryngeals and of the velar *ħ* (for the glottal stop and *h*, see § 21.1.3.1, § 21.1.3.2), and by the palatalization of the dentals, of the sibilants and of *l*, *n*.¹

21.1. The sounds

21.1.1. *Phoneme inventory*

There are 30 consonants and 7 vowels in Amharic. For the consonants, see Table 21-1.

For the other rounded consonants, see § 21.3.

The sounds *s*, *š*, *z*, *ž*, *ṣ* are also called "sibilants."

21.1.2. *Pronunciation of the Amharic consonants*

21.1.2.1. *Comparison with English*

Some Amharic sounds are the same, or nearly the same, as the English sounds, whereas others have no counterpart in English.

The sounds that are approximately the same as in English are: *b*, *p*, *m*, *f*, *w*, *s* (as in 'sun'), *z*, *y*, *g* (as in 'go'), *k*, and *h*.

In addition to these, however, some sounds that are the same or nearly the same as English sounds are rendered in the phonetic script by special symbols. These are:

1. ABBREVIATIONS: Amh., Amharic; Ar., Arabic; C, consonant; G., Ge'ez; V, vowel.

Table 21-1. Amharic Consonants

		Labials	Dentals	Palatals	Velars	Laryngeals
Stops	voiceless	p	t	č	k	(ʔ)
	voiced	b	d	ğ	g	
	glottalized	p̥	t̥	č̥	q	
	rounded				k ^w , g ^w , q ^w	
Spirants	voiceless	f	s	š		h
	voiced		z	ž		
	glottalized		s̥			
	rounded					h ^w
Nasals	voiced	m	n	ñ		
Liquids	voiced		l, r			
Semivowels		w		y		

č Ṫ corresponding to *ch* in ‘church’

š Ṣ corresponding to *sh* in ‘shoe’

ğ Ǧ corresponding to *j* in ‘joke’

ž Ț corresponding to *s* in ‘pleasure’

ñ Ȧ corresponding to *ni* in ‘onion’

The Amharic sounds that are found in English but are pronounced somewhat differently are *d* and *t*. In Amharic they are of the dental type, that is, the tip of the tongue touches the upper part of the teeth, whereas in English they are of the alveolar type, that is, the tip of the tongue touches the alveolar ridge above the teeth.

The Amharic *l* is of the “light” type as in French, German, or Spanish. That is, the tip of the tongue touches the alveolar ridge and the middle of the tongue is closer to the roof of the mouth than it is for the English *l*. The Amharic *l* is more like the English *l* at the beginning of the word (as in ‘light’) rather than the English *l* that in other positions is of the “dark” type, that is, the back of the tongue is depressed.

The Amharic *r* is a flap as in Spanish or in Italian,² the tip of the tongue striking the gum ridge behind the upper teeth.

The glottal stop ʔ corresponds to the pronunciation of English ‘uh-uh’ used as a negation or ‘oh-oh’, as an expression of surprise or warning.

2. The geminated *r* is a trill.

21.1.2.2. Consonants without English parallels

The sounds that are characteristic of Amharic and are not found in English are **ṡ** *t*, **ṣ** *ç*, **ṣ** *p*, **ṣ** *q*, and **ṣ** *s*.³ These sounds are called “glottalized” or “ejectives.” These glottalized sounds have their counterpart in nonglottalized sounds as follows:

- glottalized *t* — nonglottalized *t*
- glottalized *q*⁴ — nonglottalized *k*
- glottalized *p* — nonglottalized *p*
- glottalized *ç* — nonglottalized *ç*
- glottalized *s* — nonglottalized *s*

The difference between glottalized and nonglottalized consonants is best described as follows:

1. Nonglottalized consonants are aspirated as in English, whereas glottalized consonants are not.

2. Nonglottalized consonants are pronounced as described above, whereas in pronouncing glottalized consonants the stream of air coming from the lungs is shut off by closure of the glottis. The air above it is then forced out through a stricture somewhere along the vocal organ. The stricture is at the lips for **ṣ** *p*, at the teeth for **ṡ** *t*, **ṣ** *s*, at the palate for **ṣ** *ç*, and at the velum for **ṣ** *q*.

21.1.3. Observations on the consonants

21.1.3.1. *ʔ*

ḥ *ʔ* may be pronounced in initial pre-vocalic position, and in medial position between vowels, as is *antä* or *ʔantä* ‘you’, *säʔat* and *säat* ‘hour’, *bəʔər* and *bəər* ‘pen’. It is also pronounced in the position VC, as in *məʔrab* ‘west’, *maʔzän* ‘angle’.

For *ʔ* pronounced regionally instead of *q*, see § 21.1.3.5.

21.1.3.2. *h*

ḥ *h* (going back to *h*, *ḥ*, *ḥ*) occurs in words borrowed from Ge‘ez: e.g., *haymanot* ‘belief’, *hayl* ‘power’, *həgg* ‘law’.

It occurs in free variation with *zero*: thus, *hagär* and *agär* ‘country’, *hassab* and *assab* ‘thought’, *hamus* and *amus* ‘Thursday’.

3. These sounds are also transcribed as *t'*, *ç'*, *p'*, *k'*, and *s'*.

4. The transcription with *k* would be more consistent, but because *q* is available in the Latin alphabet, it has been used for the glottalized *k*.

An original laryngeal is *zero* in the verb, but it appears in the noun. Thus, *azzänä* ‘be sad’, but *hazän* ‘sadness’; *allämä* ‘to dream’, but *həlm* ‘dream’; *ammämä* ‘be sick’, but *həməm* ‘disease’; *annätä* ‘build’, but *hənša* ‘building’.

In initial position it may be derived from *k*. Examples: *honä* ‘be’ (G. *konä*), *hod* ‘stomach’ (G. *käbd*), *hedä* ‘go’ (G. *kedä* ‘tread’), *hulätt* ‘two’ (G. *kəlʔe*), *hullu* ‘all’ (G. *kʷəllu*).

21.1.3.3. *d ~ r*

Medial *d* in contact with a following consonant may occasionally be pronounced like *r*: thus, *qädmo* ‘before’ pronounced *qärmo*; *hedku* ‘I went’ pronounced *herku*; *gud näw* ‘it is extraordinary’ pronounced *gur näw*.

Variants of *d* and *r* occur in writing: *gədgədda* and *gərgədda* (by dissimilation) ‘wooden wall’; *gudgʷad* and *gurgʷad* (by dissimilation) ‘pit, hole’.

An original *m* of Ge‘ez (passing though *nr*) may become *nd* in Amharic. Thus, *qärm* ‘horn’ > **qänr* > Amharic *qänd*; *šərnay* ‘wheat’ > **šənray* > Amh. *sənde*; *qərnəb* ‘eyebrow’ > **qənrəb* > Amh. *qəndəb*; *kʷərnaʕ* ‘cubit’ > **kʷənraʕ* > *kənd*.

21.1.3.4. *ʃ*

The glottalized **ʌ** *ʃ* normally became **Ṣ** *t*, though it is still preserved in the region of Gojjam and Gondar; in the dialect of Wollo only **Ṣ** *t* is used.⁵ Whenever palatalization occurs (see § 21.5), *ʃ* is palatalized into *ç* as is the case of *t* (see § 21.5.1).

ʃ is preserved in expressions dealing with religion and culture. Thus, *šomä* (also *tomä*) ‘fast’, *šadəq* ‘just’, *māšhaf* ‘book’, *šəhay* ‘sun’, *šəgge* ‘flour’, and others. In the rural areas these words are pronounced with *t*: thus, *təhay*, *tay* ‘sun’, *mätəf* ‘book’, *tədq-an* ‘righteous, pious’ (pl.), *təgge* ‘flower’, and so on.

21.1.3.5. *q*

The glottal *q* has a regional variant of a glottal stop: thus, *bäʔlo* ‘mule’ for *bäqlo*; *baʔela* ‘bean’ for *baqela*; *mäʔäs* ‘scissors’ for *mäqäs*; *təlləʔ* ‘big’ for *təlləq*. There are also spelling variants with **Ḥ** used for **Ḷ**: thus, **Ḥ-C-Ḥ** *urʔ* ‘uncircumcised’ for **Ḷ-C-Ḷ** *qurʔ*; **Ḥ-C-Ḥ-Ḥ** (?) *arfaffa* ‘tall, clumsy and ungainly’ for **Ḷ-C-Ḷ-Ḷ** *qärfaffa*; **Ḥ-Ḥ** *buʔa* ‘hernia’ for **Ḷ-Ḷ** *buqa*.

5. By hypercorrectness some words are written with **ʌ** : **Ṣ** instead of **Ṣ** : **Ṣ**: e.g., **ʌ-Ṣ** ‘quarrel’ instead of **Ṣ-Ṣ**; **ṢṢ** instead of **ṢṢ**.

21.1.3.6. *č*

Medial *č* in contact with a following consonant may be pronounced like a glottal stop or *y*: e.g., *af^wačt^wall* pronounced *af^waʔt^wall* or *af^wayt^wall* ‘he has whistled’; *tägačt^wall* also pronounced *tägayt^wall* ‘it collided’; *näčč šənkurt*, also pronounced *näʔ šənkurt* ‘garlic’.

21.1.3.7. *p*

p appears in Greek loanwords borrowed through Ge‘ez: *pappas* ‘bishop’, *ṭäräppeza* (also pronounced *ṭäräbbeza*) ‘table’. For *ityoppəya* ‘Ethiopia’ pronounced *ṭobbəya*, see § 21.10.3.

p appears in modern loanwords: *polis* ‘police’, *posta* ‘mail’, *parlamänt* ‘parliament’, *polätika* ‘politics’. In the rural areas this letter is pronounced *b* (occasionally *f*): thus, *bosta*, *bolätika*, *bolis* (also *folis*).

v appears in modern loanwords: e.g., *viza* ‘visa’, *vino* (also *bino*) ‘wine’.

21.2. Phonemic status of consonants

21.2.1. *Minimal pairs*

All consonants except *ʔ* are phonemes, that is, the unit of a sound when contrasted with another unit within the same environment brings about another meaning. Only minimal pairs of glottalized versus nonglottalized, of labiovelars versus plain velars, and of palatals versus nonpalatals are given below.

21.2.1.1. *Glottalized versus nonglottalized*

- t-ṭ*: *tälla* ‘become wormy’: *ṭälla* ‘hate’; *mätta* ‘hit’: *mätṭa* ‘come’
s-š: *sällälä* ‘become paralyzed’: *šällälä* ‘become clear’
č-č̣: *mäči* ‘one who hits’: *mäč̣i* ‘one who comes’
k-q: *kämmämä* ‘trim’: *qämmämä* ‘season, spice’

21.2.1.2. *Labiovelars versus plain velars*

- g-g^w*: *gäddälä* ‘kill’: *g^wäddälä* ‘be incomplete, miss’
k-k^w: *kässäsä* ‘accuse’: *k^wässäsä* ‘become meager’
q-q^w: *qäṭṭärä* ‘hire’: *q^wäṭṭärä* ‘count’

21.2.1.3. *Palatals versus nonpalatals*

- t-č*: *tärättärä* ‘unravel’: *čäräččärä* ‘sell at retail’
ṭ-č̣: *ṭarä* ‘try very hard’: *č̣arä* ‘scrape’
s-š: *sällälä* ‘spy’: *šällälä* ‘sing a war song’

21.2.1.4. *ž ~ ġ*

Minimal pairs of *z-ž* and of *d-ġ* are more difficult to establish, as *ž* and *ġ* normally alternate. As a result, a verb such as *ġämmärä* ‘begin’ might be considered a member of the minimal pair of *zämmärä* ‘sing’ or *dämmärä* ‘sum up’.

21.3. Rounding

21.3.1. *General*

Nearly all the consonants may be pronounced with a slight rounding of the lips, which is expressed by the phonetic symbol of a raised ^w, as in *m^w*, *š^w*, *k^w*, *q^w*, *f^w*, and so on. Only the velars, however, that is, *g*, *k*, *q*, and *h* (representing *h^w*),⁶ have a complete set of graphic symbols for rounding, as, for instance, *q^wä*, *q^wi*, *q^wa*, *q^we*, *q^wə*; whereas the other consonants have a set only for the vowel *a*, as in *ḡ^wa*, *ḡ^wa*, *ḡ^wa*. For the labiovelars as phonemes, see § 21.2.1.2.

21.3.2. *Loss of labialization in velars*

A labiovelar in any position followed by *ä* may become a plain velar followed by the labial round vowel *o*. Thus, *q^wäṭṭärä* or *qoṭṭärä* ‘count’; *g^wärräfä* or *gorräfä* ‘flood, overflow’; *k^wännänä* or *konnänä* ‘condemn’. As 2nd or 3rd radical: *däqq^wäsä* or *däqqosä* ‘grind to a fine powder’; *dägg^wäsä* or *däggosä* ‘make designs on a book cover with a metal tool’; *māk^wännən* or *mākonnən* ‘governor, noble’.

A labiovelar in any position followed by *ə* usually becomes a plain velar followed by the labial rounded vowel *u*. Thus *q^wərs* or *qurs* ‘breakfast’; *g^wəd-day* or *gudday* ‘affair’; *k^wəllab* or *kullab* ‘hook of iron’.

21.3.3. *Loss of labialization in other consonants*

As stated above, the consonants other than the velars may also be pronounced rounded, the rounding being expressed in writing only when the consonants are followed by the vowel *a*, as in *ḡ^wa*. It so happens that the labials *b*, *m*, *f*, the dental *d*, and the palatal *š* may be followed by the vowel *o* in the verb or in the noun: thus, *molla* (and *mälla*) ‘be full’, *monnänä* ‘become simple, foolish’, *bonnänä* ‘rise (dust)’, *foqäffoqä* ‘be fat’, *šolläkä* ‘slip out’, *domäddomä* ‘be dull’, and others. Although the round labial vowel *o* following a labial may be owing to the preceding labial by assimilation, the vowel *o* preceded by *d*, *š* cannot be explained by assimilation to the preced-

6. *h^wa* is written only *ḡ*, originally *h^wa*, pronounced *h^wa*.

ing consonant. Neither in writing nor in pronunciation is there any indication that the above-mentioned consonants may be variants for original rounded consonants, that is, *b^wä*, *m^wä*, *s^wä*, *d^wä*, and so on, but the imperfect or jussive of the above-mentioned verbs is a clear indication that these consonants were rounded. Indeed, the jussive of these verbs is pronounced rounded: thus, *yəm^wla*, *yəb^wnän*, *yəš^wläk*. The imperfect of the quadriradicals is likewise pronounced *yəboräbb^wəral* or *yəboräbburall*, *yədomädd^wəmall* or *yədomäddumall*.⁷

As Amharic has no way to express in writing the rounded consonant + ə for consonants other than the labiovelars, it uses the spelling of the plain consonant + *u*: thus, *yəmula*, *yədomdum*, *yəšuläk*, *yəboräbburall*, *yədomäddumall*, and so on.

21.3.4. Loss of rounding

At times a labiovelar or any rounded consonant may lose its rounding; it is transferred to the preceding consonant which is then followed by the round vowel *o* or *u*, depending on the vowel of the labiovelar or on the vowel of the consonant preceding the labiovelar.

Examples for *o* transferred to the preceding consonant: *bäqq^wält*, *bäqqolt* > *boqqält* ‘sprout’; *nägg^wädä* > *noggädä* ‘thunder’; *faggorä* > *foggärä* ‘paw the ground’; *mänäk^wse* > *mänokse* ‘monk’; *mäg^wzit* > *mogzit* ‘nursemaid’.

Examples for *u* transferred to the preceding consonant: *ləg^wam* > *lugam* ‘bridle’; original *yəš^wläk* written *yəšuläk* is also written and pronounced *yuš-läk*.

The rounding may also be transferred to the following consonant: thus, *ak^wsəm* ‘Axum’ became *aksum*.

21.4. Gemination

21.4.1. General

All the consonants except *ʔ* and *h* may occur in either a geminated or a non-geminated form. Gemination is most conveniently described as lengthening of the consonant.

Gemination permeates every aspect of the morphology; hardly a sentence lacks a geminated consonant. There may be as many as five geminated consonants in one word: *lämmənnəttämmänəbbät* ‘to the one in whom we have confidence’.

7. Note that in the West Gurage cluster all the labials and velars may be rounded.

Words with two, three, or four geminated consonants are frequently seen in the language: *wättaddär* ‘soldier’, *ṭäffabbəññ* ‘it disappeared to my detriment’, *bämməttədärsubbät* ‘in that you will arrive at it’.

21.4.2. *Position*

Gemination occurs only in medial or final position.

Examples for medial position: *läbbäsä* ‘wear (a garment)’, *allä* ‘there is’, *yəmməttall* ‘he is hit’.

For final position, note the contrast between the nongeminated and the geminated consonants in *ləḡ* ‘child’ as against *əḡḡ* ‘hand’; *bäḡ* ‘sheep’ as against *däḡḡ* ‘kind’; *ḡəb* ‘hyena’ as against *ləbb* ‘heart’.

21.4.3. *Final*

The gemination of a final consonant which at times may hardly be audible becomes evident when the consonant is followed by a vowel: thus *wändəm(m)* ‘brother’, but *wändəmmə* ‘my brother’. The gemination of the final *l* in the compound imperfect as in *yənäḡrall* is more audible when followed by a vowel as in *yənäḡrallu*.

21.4.4. *Phonemic status*

Gemination is phonemic in Amharic. The following minimal pairs illustrate this feature: *gāna* ‘still’ : *gānna* ‘Christmas’; *alä* ‘he said’ : *allä* ‘there is’; *šəfta* ‘outlaw’ : *šəffəta* ‘rash’; *fəraš* ‘mattress’ : *fərräš* ‘remains, ruin’; *yəmətall* ‘he hits’ : *yəmməttall* ‘he is hit’.

21.4.5. *Assimilation*

Gemination also results from the assimilation of one consonant to another: thus, *yəssäbbär* ‘it is broken’, from **yətsäbbär*; *əssu* ‘he’ and *ərsu*; *arrässamm* ‘he didn’t forget’, from *alrässamm*.

21.4.6. *Lexical and morphological*

Gemination is either a lexical or a morphological feature. Gemination as a lexical feature cannot be predicted. Thus, for instance, **ዋና** may be read *wan-na* meaning ‘main, principal’ or *wana* meaning ‘swimming’.

As a morphological feature gemination is more predictable in the verb than in the noun. Thus, the 2nd radical of a triradical verb or the 3rd radical of a quadriradical verb is always geminated in the perfect (*säbbärä*, *mäsäkkärä*). Likewise, the gemination or the nongemination of the 2nd radi-

cal in the imperfect is predictable when the type of the verb is known. Indeed, in type A the 2nd radical is not geminated (*yəsäbr*), whereas in a type-B verb the 2nd radical is geminated (*yəfälləg*).

The situation with the noun is less clear. In a primary noun there is no way of knowing whether any one of the radicals is geminated or not. Thus, ሸረረት could be read *šärärit*, *šärrärit*, *šärärrit*, or even *šärrärrit*.

In some patterns of the derived nouns the gemination of a radical is predictable, whereas in other patterns this is not. Thus, whereas in the pattern ቅትትል the 3rd letter is geminated regardless of whether the noun derives from a type-A or a type-B verb, as in *nəgəggər* ‘speech’ (type A), *ləwəwwət* ‘exchange’ (type B)—in the pattern ቀትል *qät(t)äl* of type A, *sänäf* ‘lazy’ has no gemination, but *addäg* (likewise of type A) ‘one who has grown’ has gemination.

For the dissimilation of gemination, see § 21.10.2.

21.5. Palatalization

21.5.1. General

As noted above (§ 21.2.1.3), the palatals are phonemic. The process of palatalization is also a phonetic feature. Indeed, the dentals *d*, *t*, *t̪*, *l*, and *n*, and the sibilants *s*, *z*, and *ʃ* are palatalized when followed by the vowel *-i*, *-e*, or *-iya*. The resulting sounds are: *di* > *ǵ*, *ti* > *č*, *t̪i* (*ʃi*) > *č̣*, *li* > *y*, *ni* > *ñ*, *si* > *š*, *zi* > *ž*.

Palatalization occurs in verbs and in verbals, that is, the active participle and the instrumental, nouns that are derived from the verb.

21.5.2. In verbs

The forms that end in *-i* are the imperfect, jussive, and imperative, the singular 2nd person feminine, and the active participle; the form that ends in *-e* is the gerund, 1st person singular; and the form that ends in *-iy(a)* is the verbal noun.

The vowel *i* following the above-mentioned consonants may either be kept with the palatal or, more usually, be absorbed. Thus,

wəsäǵ or *wəsäǵi* ‘take!’ from *wässädä*

kəfäč or *kəfäči* ‘open!’ from *käffätä*

bələč or *bələči* ‘exceed!’ from *bällätä*

kəfäy or *kəfäyi* ‘pay!’ from *käffälä*

lämmən(ñ) or *lämmənñi* ‘beg!’ from *lämmänä*

mälləš or *mälləši* ‘return!’ from *mälläsä*

märrəž or *märrəži* ‘poison!’ from *märräzä*

In the active participle the final *-i* is normally absorbed in the palatal: *araš* (from *arräsä*) ‘he who plows’. In some isolated participles, however, the final dental or sibilant is not palatalized. The participles *därasī* ‘author, writer’ and perhaps *anaṭī* ‘carpenter’ may be borrowed from Ge‘ez, but the same is not true of *adli* ‘one who is partial’ (from *a-dälla*).

21.5.3. *In the gerund*

The expected vowel *e* in the gerund, 1st singular, may be pronounced and written *ä* after the palatals: thus, *mälləšše* or *mälləššä* ‘I having returned’; *käfəčče* or *käfəčcä* ‘I having opened’ (see § 21.21.2).

21.5.4. *In the instrumental*

The instrumental is *mälqām-īya* or *mälqām-ya*. The sequence *-īya* or *-ya* causes palatalization of the sibilants and dentals and is absorbed in the palatal. Examples: *mäkfäča* ‘key’ from **mäkfätiya*; *maräša* ‘plow’ from **maräsiya*. In a verb ending in *I*, *-liya* becomes *-ya*; thus, *mänqäya* ‘instrument for pulling out’, from *mänqäliya*.

21.5.5. *In nouns*

The final vowel *e* in a noun does not cause palatalization of a dental or a sibilant: thus, *bet-e* ‘my house’. Some nominal patterns, however, bring about palatalization, such as the patterns *qättil*, *qəttali*, *qərtali*, *qətləya*, *qətləyya*, and the pattern ending in *-yyo*.

Examples for *qättil*: *aččir* ‘short’ (from *aṭṭärä*), *räžžim*, *rägğim* ‘long’ (from *räzzämä*), *qäččin* ‘thin’ (from *qäṭṭänä*), but without palatalization in *addis* ‘new’, *qällil* ‘light’.

Examples for *qəttali*, *qəṛəttali*: *məllaš* ‘answer’ (from *mälläsä*), *fərraš* ‘ruin’ (from *färräsä*), *gʷədday*, ‘incomplete, wanting’ (from *gʷäddälä*), *gəžži* or *gəžž* ‘purchase’ (from *gäzza*), *kəṭəkkač* ‘chopped up’ (from *kätäkkätä*).

Examples for *qətləya*, *qətləyya*: *mərča* ‘choice’ (from *märräṭä*), *wərgə* ‘miscarriage’ (from *wärrädä*), *əṛša* ‘field’ (from *arräsä*), *ərməgğə* ‘step, pace’ (from *tärammädä*), *gədəyya* ‘killing’ (from *gäddälä*).

Examples for nouns ending in *-yyo*: *mälašo* ‘salt given to cattle to lick’ (from *lasä*), *mäqqäčo* ‘fine’ (from *qäṭṭa* ‘punish’).

21.5.6. *In other parts of speech*

Palatalization also occurs in other parts of speech. Thus, *anči* ‘you’ (fem.) from *anti*; *mäčä* ‘when?’, from **matay*.

21.5.7. Regional palatalization

Outside the special nominal patterns there are also regional (Gojjam, Gondar, Wollo) occurrences of palatalization of consonants followed by the vowel *i*, *e*, or *ə*. Thus,

- k* > *č*: *kidan* ‘pact’ > *čidan*; *ənki* ‘take!’ > *ənči*
q > *č*: *qes* ‘priest’ > *čes*; *duqet* ‘flour’ > *dučät*
g > *ğ*: *gize* ‘time’ > *ğəze* (Gojjam); *aroge* ‘old’ > *arogħe*
d > *ğ*: *dəngula* ‘stallion’ > *ğəngula*
t > *č*: *ətəge* ‘head of the Ethiopian monks’ > *əčəge*

21.5.8. Depalatalization

By an inverse process *ğ* may become *g* in various regions: *əğğeta* and *eggeta* ‘sleeve’; *ğärba* and *gerba* ‘back’; *ğoro* and *gero* ‘ear’.

21.6. Weakening of consonants

21.6.1. *ğ* ~ *y*

The palatal *ğ* may occasionally become *y*. Examples: *dawğa* and *dawya* ‘mat woven of reeds’, *täzägağtäwall* and *täzägaytäwall* ‘they are ready’.

By an inverse process *y* may become *ğ*: thus, *əyubbəññ* and *əğubbəññ*, formula for making someone one’s attorney; *moyale* and *moğale* ‘a small flea’.

21.6.2. *w* ~ *y*

The labial *w* may become *y*. Examples: *sārawit* and *sārayət* ‘army’, *dawit* and *dayət* ‘Psalter’, *māstawät* and *māstayät* ‘mirror’.

The labial *b* passing through the spirantized *ḅ* may become *y*: thus, *ṭäbib* and *ṭäyəb* ‘smith’, *zäbib* and *zäyəb* ‘raisin’.

21.7. Spirantization

21.7.1. *b* ~ *ḅ*

A nongeminated intervocalic *b* is spirantized and pronounced *ḅ*: thus, *səḅär* ‘break!’. The city of *addis abäba* is pronounced *addis aḅäḅa* and, by haplology, *addisaḅa*.

The alternate forms of *wäbo* and *wäwo* ‘rear guard’ (perhaps by assimilation), of *səlbabot* and *səlbawät* ‘film on the surface of hot milk’ (perhaps by dissimilation of *b-b* > *b-w*; see § 21.10.1.1), and of *ab"ara* and *awwara* ‘dust’ are also examples of spirantization.

Historically Amh. *sāw* ‘man, person’ goes back to Ge‘ez *sābəʔ*, and *ṭwat* ‘morning’ goes back to Ge‘ez *ṣəbḥat*.

An inverse process of *w* becoming *b* occurs in *abol* ‘first infusion of coffee’ from Arabic *ʔawwal* ‘first’.

21.7.2. *k ~ h*

A nongeminated intervocalic *k* may likewise regionally become a spirantized *h*. This is the case of *nəka* ‘touch!’ becoming *nəha*; *yənəkall* ‘he touches’ becoming *yənāhall*; *kābbādā* ‘be heavy’, but *yāhābbādā* ‘that which is heavy’ in the dialect of Mānz. For the alternation of *k:h*, see § 21.8.2.1.

21.8. Sporadic alternation

This feature refers to the unpredictable substitution of one phoneme for another. It occurs with labials, velars, liquids, and palatals versus nonpalatals. Although some occurrences of sporadic alternation might be explained through regional or colloquial usage, in other occurrences assimilation might have played a role. Sporadic alternation also occurs in loanwords.

21.8.1. *Alternation of labials*

21.8.1.1. *m-b, b-m*

m instead of *b*: *məntä* ‘for the sake of’ instead of **bəntä* (from Ge‘ez *bäʔəntä*).

Arabic loanwords: *qalib* and *qalim* ‘mold’ (Ar. *qālib*); *ṭābāŋga* and *ṭāmāŋga* ‘rifle’ (Ar. *ṭabāŋga*, of Turkish origin).

m instead of *f*: *əffuññit* and *əmmuññit* ‘viper’.

b instead of *m*: *mahtāb*, *matāb* ‘seal’ instead of *mahtām*; *zābān* ‘time’ instead of *zāmān*.

Arabic loanwords: *bilqaṭ* ‘perfume bottle’ (Ar. *milqaṭ*); *əlləmat* and *əlləbat* ‘bookmark’ (Ar. *ṣilāma*).

b instead of *f*: *nāfs* and *nābs* ‘soul’; *bulad* ‘flint, steel’, from Arabic *fūlād*.

f instead of *b*: *kʷārābta* and *kʷārāfta* ‘hill’.

In loanwords: *p* instead of *b*: *lamba* and *lampa* ‘lamp’ (from Arabic *lamba* or perhaps from Italian *lampa*).

b instead of *p*: *pakko* and *bakko* ‘pack (of cigarettes)’ from Italian *pacco*; *polātika* and *bolātika* ‘politics’.

b instead of *p*: *zopi* and *zobi* ‘ebony’.

f instead of *p*: *polis*, *bolis*, and *folis* ‘police’.

21.8.2. Alternation of velars

21.8.2.1. *k-h, h-k*

Amharic *h* goes back to *h*, *ḥ*, and *ḥ̣*. In fact, these letters are used in the orthography, but they are pronounced as *h* (see § 21.1.3.2). Examples: *haket* ‘indolence’ (Ge‘ez *hakäyā* ‘be indolent’); *ḥəyaw* ‘immortal’ (G. *ḥəyaw*); *ḥəlli-na* ‘conscience’ (G. *ḥəlli-na*).

The verbal suffixes of the perfect *-hu* (1st person), *-h* (2nd person) go back to *-ku*, *-k*, as in *näggärhu*, *näggärh* ‘I said’, from *näggärku*; *näggärh* ‘you said’, from *näggärk*. See also § 21.8.2.1.1.

21.8.2.1.1. *h < k*

Some lexemes with Amharic *h* go back to *k*. Examples: *hullu* ‘all’ (G. *kʷəllu*); *hulätt* ‘two’ (G. *kəllə*); *hʷala* ‘after’ (G. *käwäla* ‘behind, back’); *honä* ‘be’ (G. *konä*); *hod* ‘stomach’ (G. *käbd*); *hedä* ‘go’ (G. *kedä* ‘tread’). See also § 21.8.2.1.

21.8.2.1.2. *k ~ h*

In Amharic itself there are several lexemes with an etymological *k* which have variants with *h*. Thus: *kʷämätṭätä* ‘be sour, acid’ and *homätṭätä*; *kudade* ‘Lent’ and *hudade*; *məknəyat* ‘cause, reason’ and *məhnəyat*; *šəkona* ‘sole of the foot’ and *šəhona*; *əšok* ‘thorn’ and *əšoh*.

It also happens that a false reconstruction of an original *h* becomes *k*, as in *homär* (from Ar. *ḥumar*) and *komär* ‘tamarind’; *kamsa* ‘fifty’ for the existing *hamsa*.

21.8.2.2. *k-q*

karra and *qarra* ‘steel knife’; *əmbʷaka* and *əmbʷaq* ‘coward’.

21.8.2.3. *g-k, k-g*

gʷärrätä and *kʷärrätä* ‘make a pile of stones’; *gäwgawwa* and *käwkawwa* (also *qäwqawwa*) ‘who wanders about aimlessly’; *kufeta* ‘skullcap’ (from Arabic *kūfiya*) and *gufta* ‘scarf’.

21.8.3. Alternation of liquids

21.8.3.1. *l ~ n*

l instead of *n*: *mänäkʷse* and *mäläkʷse* ‘monk’; *ḡanmeda* and *ḡalmeda* ‘race-course in Addis Ababa’; *zäläb* ‘tail’, from Arabic *ḡanab*.

n instead of *l*: *qəmal ḡər* and *qəman ḡər* ‘licelike body parasite’; *läqämt* and *näqämt*, city in the Wollega region.

21.8.3.2. *n ~ m*

n instead of *m*: *šāmqaqqo* and *šanqaqqo* ‘snare for trapping animals’.

m instead of *n*: *konso* and *komso* ‘a kind of handmade cotton blanket’.

From Turkish-Arabic: *barneta* and *barmeta* ‘European-style hat’.

21.8.3.3. *r ~ l*

r instead of *l*: *gələmbiṭ* and *gərəmbiṭ* ‘opposite, contrary’ (from *gäläbbätä* ‘turn upside down’).

Note that in some occurrences assimilation or dissimilation may have brought about the alternance of the liquids.

21.8.4. *Alternation of sibilants*

21.8.4.1. *s-š*

sāk^wāna and *šāk^wāna* ‘hoof, foot’; *səntər* and *šəntər* ‘splinter’; *mist* and *mišt* ‘wife’; *wāgessa* and *wāgešša* ‘traditional medical practitioner’.

From Arabic: *sāmbuq* and *šāmbuq* ‘small sailing vessel with sail’; *sārṭan* and *šārṭan* ‘crab’; *qərṭas* and *qərtaš* ‘piece of paper’.

21.8.4.2. *z-ž-(ǧ)*

(*a*)*mānāzzāhā* and (*a*)*mānāžžāhā* ‘chew’; *mize* and *mənǧe* ‘best man’ (for inserted *n*, see § 21.12); *zəgra* and *žəgra* ‘guinea hen’.

21.8.5. *Alternation of palatals*

21.8.5.1. *š ~ č ~ ǧ*

š:č: *šəfal* and *čəfal* ‘eyebrow’; *čəgəñ* and *šəgəñ* ‘seedling, plant shoot’; *čəggər* and *šəggər* ‘trouble, difficulty’; (*a*)*ššānnāfā* and (*a*)*čānnāfā* ‘overcome’.

š:ǧ: *šəgəl* and *ǧəgəl* ‘gonorrhea’.

21.8.5.2. *ž ~ ǧ*

ž is in free variation with *ǧ* in any position. Initial: *žəgra* and *ǧəgra* ‘guinea hen’; *žāmmārā* and *ǧāmmārā* ‘begin’; *žoro* and *ǧoro* ‘ear’.

Medial: *bāžād* and *bāǧād* ‘abdomen’; *anžāt* and *anǧāt* ‘bowels’.

Final: *goǧǧo* and *gožžo* ‘small hut’; *g^wāmāǧǧä* and *g^wāmāžžä* ‘desire eagerly’.

21.9. Assimilation

21.9.1. Total regressive assimilation

21.9.1.1. Consonants

Total assimilation occurs when two consonants are in contact. It affects the dentals *d*, *t*, *t̥*; the liquids *l*, *r*; the sibilants *s*, *z*; and the velars *g*, *q*.

21.9.1.2. Morpheme *t*

The morpheme *t* of the imperfect, jussive, and of the verbal noun of the *tä*-stem is assimilated to any consonant with which it is in contact: thus, from *tākāffātä* ‘be opened’: imperfect *yēkkāffāt* for *yētkāffāt*, jussive *yēkkāfāt* for *yētkāfāt*, verbal noun *mākkāfāt* for *mātkāfāt*, instrumental *mākkāfāča* for *mātkāfāča*.

The derived stem *annaggärä* likewise derives from **atnaggärä*.

The liquid *l* may be assimilated to an initial *r* of the verb: thus, *arrāddamm* ‘he didn’t help’ (for *alrāddamm*), but also *alrāddamm*.

The liquid *r* is assimilated to *s* in *əssu* ‘he’, *əss* ‘a’ ‘she’, but also *ərsu*, *ərs* ‘a’.

21.9.1.3. Dentals

The glottalized *t̥* is assimilated to the following *t* in the gerund: thus, *mätto* ‘he coming’ for *mät̥to*; *wätta* ‘she going out’ for *wät̥ta* (but also *mät̥to*, *wät̥ta*; see § 21.9.2).

The dental *d* becomes assimilated to the following *t* in the gerund: thus, *nädto* ‘he driving’, pronounced *nätto*; *kädto* ‘he betraying’, pronounced *kätto*; *šädto* ‘it being clean’, pronounced *šätto*.

The dental *d* is assimilated to the following *ǧ*: thus, *təwädǧ* ‘you (fem.) love’ (from *wäddädä*) becomes *təwäǧǧ* or *təwäǧǧi*.

21.9.1.4. Sibilants

The final *s* of the morpheme *as-* becomes assimilated to the initial *z*, *š*, or *ʃ* of the verb: thus, *azzämmätä*, for *aszämmätä* ‘have someone go on campaign’; *aššällämä*, for *asšällämä* ‘have someone decorated’; *aššäyyäfä*, for *asšäyyäfä* ‘fill with disgust’.

In other occurrences, too, the succession *sš* becomes *šš*, *zš* becomes *žž*, and *zš* becomes *šš*: thus, *təkäšš* or *təkäšši*, from *təkäsš* ‘you (fem.) accuse’ (from *kässäsä*); *tažž* or *tažži* ‘you (fem.) command’ (from *azzäzä*).

21.9.1.5. Lexical items

In the frequently used verbs *ǧäddälä* ‘kill’, *ǧ* “*äddälä* ‘be missing’, *d* is assimi-

lated to the following *l*: thus, *gällo* ‘he killing’, for *gädlo*; *yəgällall* ‘he kills’, for *yəgädlall*; *g^wällo* ‘it missing’, for *g^wädlo*; *yəg^wällall*, for *yəg^wädlall*.

Likewise the frequently used *dägmo* ‘again, furthermore’ is pronounced *dämmo*.

For the frequently used *adärrägä*, see § 21.14.

21.9.1.6. *Velars*

The velars *q*, *g* become assimilated to the suffix pronoun *-k* of the 1st and 2nd singular: thus, *särräkk* ‘you stole’, for *särräqk*; *fälläkk* ‘you wished’, for *fälläqk*; *altäyyäkkumm* ‘I didn’t ask’, for *altäyyäqkumm*.

21.9.1.7. *Miscellaneous*

Isolated occurrences of assimilation: *tunča* becoming *tučča* ‘shoulder blade’; *mənčät* becoming *məččät* ‘jar, pot’; *səlsa* and *səssa* ‘sixty’; *əndet nāw?* ‘how is he?’ becoming *əndennāw*.

21.9.2. *Total progressive assimilation*

Total progressive assimilation occurs in contact of *t* of the gerund with the final *ɬ* of the verb: thus *mätto*, for *mätto* ‘he coming’; *wätta*, for *wätta* ‘she going out’ (but also *mätto*, *wätta*, for which see § 21.9.1.3).

21.9.3. *Partial assimilation*

21.9.3.1. *Contiguous assimilation*

nb may become *mb*. Thus, *šənbəra* and *šəmbəra* ‘chick-pea’; *ənbässa* and *əmbässa* ‘lion’; *ənbərt* and *əmbərt* ‘navel’; *wənbär* and *wəmbär* ‘chair’.

Likewise *nf* may become *mf*: *qərənfud* and *qərəmfud* ‘clove’; *gänfo* and *gəmfō* ‘porridge’.

21.9.3.2. *Assimilation of voicing*

Voiced-voiceless > voiceless-voiceless in *täzkar* and *täskar* ‘memorial service held on the 40th day after death’; *gugs* and *guks* ‘kind of polo’; *magsäñño* and *maksäñño* ‘Tuesday’.

21.9.3.3. *Isolated occurrences of partial assimilation*

mɬ may become *nɬ* (the labial *m* becoming the dental labial *n* in contact with the dental *ɬ*) in *amɬa* and *anɬa* ‘bring!’.

md may become *nd*: e.g., *ɬəmd* and *ɬənd* ‘pair, couple’.

tm became *tn* in *aɬənt* ‘bone’, from *ɬsm*, *ɬtm*.

ln > *nn* in *təlantənna* and *tənantənna* ‘yesterday’.

21.9.3.4. *Noncontiguous assimilation of voicing*

gutəčča and *kutəčča* ‘earrings’; *səgağğa* (from Arabic *sağğāda*) and *zəgağğa* ‘prayer rug’ (also *səgağğa*); *tämäräkk’äzä* and *tämärägg’äzä* ‘lean upon’.

Of labials *m-b* > *m-m*: *mabär* and *mamär* ‘association’.

Of liquids *l, n*: *kuntal* and *kuntan* ‘quintal’ (*n-l* > *n-n*; the original form is *qintar*, from Arabic *qintār*).

Of glottalization: *k-ʔ* > *q-ʔ*: *käräʔit* (from Arabic *ḥarīʔa*) and *qäräʔit* ‘pouch’ (from Ar. *ḥarīʔa*).

q-ʔ > *q-q* in *qärṭäm* > *qärqäm* ‘kind of plant resembling cress’; *qənṭəš* > *qən-qəš* ‘kind of sorghum’.

Of palatals: *saṅga* and *šaṅga* ‘bayonet’; *säläččä-* and *šäläččä-* ‘be bored’.

21.10. Dissimilation

21.10.1. *Contiguous dissimilation of voicing*

žg > *šg* in *mäžgär* and *mäšgär* ‘tick’; *qt* > *gt* in *aqtačča* and *agtačča* ‘direction toward something’.

21.10.1.1. *Noncontiguous dissimilation*

The labial *m* of a noun expressing an instrument becomes *w* if there is a labial *b, m*, or *f* in the noun: thus, *wämbär* ‘chair’ (from **mänbär*), *waggəmt* ‘cup used for drawing blood’ (from *aggämä* ‘draw blood by cupping’), *wät-mäd* ‘snare, trap’, *wäqlämt* and *mäqlämt* ‘small knife’, *wäsfe* ‘awl’, *wänäf* ‘bellows’, *wänṭäft* ‘strainer’, *wänfit* ‘sieve’, *wäfčo* ‘grindstone’, *wäsfänṭər* ‘springy stick which is made into a snare’ (from *fänäṭṭärä* ‘spring out’).

m-m may become *m-b*, as in *mahtäb*, *matäb* ‘seal’ instead of *mahtäm*; *mäharäm*, *mäharäb* ‘handkerchief’, from Arabic *maḥrama*.

Of two labials: *fälässäfä* and *fälässämä* ‘philosophize, discover, invent’; *anboləp* and *anboləkk* ‘envelope’.

n-n > *l-n*, or *n-l*: *nišan* ‘medal’ (from Arabic *nišān*) became *lišan*; *ganen* ‘demon’ became *ganel*.

21.10.1.2. *Of voicing*

gubbəta and *qubbəta* ‘hill, hillock’; *sənag* and *sənaq* ‘palate’.

Of glottalization (a voiceless glottalized becomes voiceless nonglottalized): *qärqäha* and *qärkäha* ‘bamboo’; *qawṭ* and *qawt* ‘kind of tree used for plow beams’.

Of glottalization and voicing (a voiceless glottalized becomes voiced

nonglottalized): *qəŋtar* and *gəŋtar* ‘quintal’; *ṭārəppeza* and *ṭārəbbeza* ‘table’; *wäqqäṭä* and *wäggäṭä* ‘pound’.

21.10.2. Dissimilation of gemination

The intensive action of type A of the composite verbs is expressed by *kəffətt adärrägä* ‘open completely and suddenly’, with gemination of the 2nd radical. In the verbs with identical second and third radical (“1.2.2”), however, the form is *wədədd alä*, with a nongeminated 2nd radical. The identity of the last two radicals brings about the dissimilation of gemination and, as a result, the 2nd radical is not geminated. The situation is the same in type-B verbs in which the attenuative action is expressed by the gemination of the 2nd radical: thus, *läzzäbb alä* ‘be somewhat soft’. In 1.2.2 verbs, again by the dissimilation of gemination, the 2nd radical is simple: thus, *läṭäṭṭ alä* ‘stretch somewhat’.

21.10.3. Reciprocal assimilation and dissimilation

qərṭas ‘piece of paper’ > **kərṭas* (by dissimilation) > *kərṭas* (by assimilation); *ṭənčäl* ‘hare, rabbit’ > **čənčäl* (by assimilation) > *čəntäl* (by dissimilation); *ityoppəya* > **ityoppəya* (by assimilation) > *ṭobbəya* (by dissimilation); *čäqäččäqä*, and *čäkäččäkä* ‘pester’.

21.11. Metathesis

21.11.1. Liquids and sibilants

Because of the relatively small number of occurrences with metathesis it is difficult to explain the reason for this feature. In the majority of occurrences the liquids *l*, *n*, *r* are involved.

With *r*: *səfra*, *sərfa* ‘place’; *māṭrəbiya*, *māṭəbiya* ‘axe’; *kābāro*, *kārābo* ‘drum’; *bärtä qaqa*n, *bātrə qaqa*n ‘iron bar’; *qərṭəndo*, *qəṭrindo* ‘kind of plant’; *gärräzä*, *gäzzärä* ‘circumcize’.

From Arabic: *märbuṭ*, *mābruṭ* ‘kind of rifle’; *mänäṭṭär*, *märäṭṭän* ‘telescope’; *kəbrit*, *kərbit* ‘matches’.

With *n*: *qəbanug*, *qənabug* ‘oil from the *nug*-seed’; *qəntəš*, *ṭənqəš* ‘kind of sorghum’; *šännäṭä*, *näššäṭä* ‘arouse great enthusiasm’.

With *l*: *sälen*, *sänel* ‘palm mat’; *bäläṭṭägä*, *bäṭällägä* ‘become rich’.

With the sibilants *s* and *z*: *maksäñño*, *maskäñño* ‘Tuesday’; *betäksiyān*, *be-täksiyān* ‘church’; *sämmäṭä*, *säṭṭämä* ‘get drowned, sink’.

With *z*: *əzger*, *əzger* ‘God’ (but also with *r* in the root); *hullägize*, *hulläzge* ‘always’ (but also with *l* in the root).

21.11.2. Metathesis in Amharic roots in relation to Ge'ez roots

Amh. *läwwätä* 'change', Ge'ez *wälätä*; Amharic *annäsä* 'be little', Ge'ez *nəʔəsä*; Amharic *təlant* 'yesterday', from *təlam-t*, Ge'ez *təmal-əm*; Amharic *awre* 'wild animal', Ge'ez *arwe*.

21.12. Insertion of *n*, *r*

21.12.1. With quadriradicals

The dental-nasal *n* is occasionally intercalated between two radicals. This process of nasalization occurs most frequently when the root contains a velar *g*, *q*, or *k*.

With a velar in quadriradicals: *gärägäre* and *gärängäre* 'resinous bush'; *gərgərit* and *gərəngərit* 'tying the hands behind the back'; *gədgədi* and *gədəngədi* 'kind of hawk'; *darəgot* and *darəngot* 'a morsel of bread'; *balägara* and *balängara* 'opponent'. In the preceding examples *n* is placed before the velar.

Placed after the velar in *duqduq* and *duqənduq* 'dung worm'.

With other consonants: *dəbəlbal* 'round', and *dənəlbal* 'sphere, ball'; *ğəlağəl* and *ğəlaŋğəl* 'imbecile'; *ğələnbiṭ* 'contrary, opposite', from *ğäləbbätä* 'turn upside down'.

21.12.2. With biradicals and triradicals

Involving a velar: *qug* and *qung* 'snapper of a whip'; *roqe* and *ronqe* 'marsh'.

With other consonants: *däf* and *dānf* 'threshold'; *mize* and *minğe* 'best man'; *əntəff alä* 'spit' (cf. *əttəff alä*, *täffa* 'spit'); *qəfəd* and *qəfənd* 'pus in the eye owing to inflammation'; *wənçəf* 'sling' from *wäççäfä* 'hurl with a sling' (cf. Ge'ez *modäf* 'sling', from **mäwdäf*, without *n*); *qäbät* and *qänbät* 'bud'; *gäzzärä* 'circumcise' and *gänäzzärä* 'cut'.

In nouns with a final *-t*: *mağrat* and *manžərat* 'nape of the neck'; *säräbät* and *säränbät* 'kind of bread made of chickpea flour'; *məzəllat* and *mənəzəllat* 'great-great-grandparents'.

21.12.3. Lexical item

Amharic *wänd* 'male' is derived from **wäld* becoming **wädd* by assimilation (cf. Tigrinya *wäddi*) > *wänd*, with inserted *n*.

21.12.4. Miscellaneous

A few roots with a velar and an inserted *n* in relation to Ge'ez or to the other Semitic languages: *anäqqäfä* 'stumble', G. *(ʔa)ṣqäfä*; *anäkkäsä* 'limp',

Ar. *ḥakasa*; *däqqä māzmur* and *dānqä māzmur* ‘student, disciple’ (G. *däqq*); *däqäl* and *dānqäl* ‘mast of a ship’.

With a consonant other than a velar: *anättäsä* ‘sneeze’, Ge‘ez *ʕatäsä*; *ənčät* ‘wood’, G. *ʕəḏ*; *wānz* ‘river’ (G. *wəḥzä* ‘flow’); and ‘one’ from **ad* (Ge‘ez *ʔahadu*).

21.12.5. *r*

A few nouns have an inserted *r*: *qurqumba* and *qurquramba* ‘bottle-shaped container’; *qumbus* and *qərənbus* ‘dung maggot’.

21.13. Abbreviations and haplology

21.13.1. Abbreviations

Frequently used nouns are abbreviated: *əgzer* (also *əzger*) for *əgziabəher* ‘God’; *betäksiyan* and *betäskiyan* for *betä krəstiyan* ‘church’; *šaqa*, or *yäšaqa* for *šäläqa* ‘captain’; *blatta* for *blattengeta*, honorific title; *däğğacč* for *däğğazmačč*, honorific title; *əče* for *əčege* ‘head of the Ethiopian monks’.

wätte emphasizes the attributes of the *wättaddär* ‘soldier’; *däbte* emphasizes the attributes of the *däbtära* ‘unordained member of the clergy’; *šəme* is used by children to make fun of the *šəmagelle* ‘old man’; *šärme* is used to make fun of the *šärmuṭa* ‘prostitute’.

21.13.2. Haplology

addis aba for *addis abäba* ‘Addis Ababa’; *yäneta*, *yänta* for *yäne geta* ‘my master’, term used by traditional church students in addressing teachers; *əmmete*, *əmmäyte*, for *əmmä bete* ‘madam’, lit. ‘mother of the house’; *filäfit* and *fit läfit* ‘directly in front of’; *təlammata* for *təlant mata* ‘last night’.

21.14. Loss of consonants

The frequently used verb *adärrägä* ‘do, make’ loses the dental *d* without leaving a trace in the verb: thus, perfect *arägä* (note the nongemination of the *r*), imperfect *yarägall* for *yadärgall*, imperative *arg* for *adrəg*, gerund *argo* for *adrəgo*, and so on.

The *d* is likewise lost in *qəmayat* for *qədmayat* ‘great-grandfather’.

The nasal-dental *n* is lost in *mədəğğä* ‘fireplace’, from *näddädä* ‘burn’.

In the frequently used verb *awwäqä* ‘know’, the consonant *w* may be omitted: thus, *yaqall* for *yawqall* ‘he knows’, *ayaqəmm* ‘he doesn’t know’ for

ayawqəmm, *yämmiyaq* ‘he who knows’ for *yämmiyawq*, *yämmayaq* ‘he who doesn’t know’ for *yämmayawq*.

The consonant *y* is lost in *adollämm*, *adällämm* for *aydollämm*, *aydällämm* ‘he is not’.

21.15. Vowels

21.15.1. Inventory

There are seven vowels in Amharic. The traditional order is: *ä*, *u*, *i*, *a*, *e*, *ə*, *o*.

Table 21-2 illustrates the position of the vowels.

Table 21-2. Amharic Vowels

	front	central	back
high	i	ə	u
mid	e	ä	o
low		a	

There is no precise correspondence in the pronunciation of Amharic and English vowels.

21.15.1.1. *i*

The vowel *i* is pronounced somewhat like the ‘ee’ in ‘feet’ but without the /y/ glide of English: *fit* ‘face’, *anči* ‘you’ (fem.).

In a final palatal the expected *i* may be omitted in certain forms in writing and in pronunciation: thus, *kəfäči* and *kəfäč* ‘open!’ (fem.); *ləbäši* and *ləbäš* ‘get dressed!’ (fem.).

21.15.1.2. *e*

The pronunciation of the vowel *e* is approximately like that of the vowel ‘a’ in ‘state’ but without the /y/ glide of English: *bet* ‘house’, *nuzaze* ‘will’. Note that in regions outside Addis Ababa all the consonants have slightly palatalized allophones when preceding the vowel *e*: thus, *b^ʲet^ʲe* ‘my house’.⁸ See also the vowel *o*, § 21.15.1.5.

8. This statement is in contradiction with some grammars which state that the vowel *e* is diphthongized as *ʲe*. In fact, even in regions where **ቤት** is pronounced *b^ʲet^ʲe*, the vowel is never diphthongized after a palatal (thus, **ደርሽ** *därəšše*, not *därəšš^ʲe*) or initially, as in **ሂሊ** *eli* ‘tortoise’.

In palatals and with *y* the expected *e* may be written and pronounced *ä*; thus, *käfäčče* and *käfäččä* ‘I opening’; *ǧelä* and *ǧälä* ‘be foolish’; *səraye* and *sərayä* ‘my job’.

21.15.1.3. *ä*

The vowel *ä* is pronounced like the sound one makes while hesitating in speaking; it is represented in writing by ‘uh’: *nägä* ‘tomorrow’, *ǧäräd* ‘servant’. It is a mid-low central vowel, halfway between [ə] and [a].

No word in Amharic begins with *ä* except *ärä* ‘then, so then, really?’

21.15.1.4. *a*

The vowel *a* is pronounced like the vowel *a* in ‘father’: *addärä* ‘spend the night’, *na* ‘come!’.

21.15.1.5. *o*

The vowel *o* is pronounced approximately like the English ‘a’ in ‘also’: *roṭä* ‘run’, *of* ‘bird’. Note that in some regions outside Addis Ababa all consonants in contact with *o* are pronounced with rounded lips, resulting in a slight *w* off-glide: thus, *män^wor* ‘existence, presence’, *m^wotä* ‘die’, *s^wost* ‘three’. See also § 21.15.1.2.

21.15.1.6. *u*

The vowel *u* is pronounced approximately like the English ‘o’ in ‘who’ but without diphthongization: *hullu* ‘all’, *tut* ‘breast’.

21.15.1.7. *ə*

The vowel *ə* is pronounced approximately like the English ‘e’ in ‘roses’: *əssu* ‘he’, *sənt* ‘how much?’

The vowel *ə* is phonemic: thus, አለቁ, the gerund of *alläqä* ‘come to an end’, is pronounced *alqo*, but አለቁ, the gerund of *a-laqä* ‘make more, increase’, is pronounced *aləqo* (likewise አርሶ, the gerund of *arräsä* ‘plow’, is pronounced *arso*, but አርሶ, the gerund of *a-rasä* ‘wet, drench’, is pronounced *arəso*); ይስማቸው is pronounced *yəsmaččäw* ‘may he listen to them’, but *yəsəmaččäw näbbär* ‘he used to kiss them’; ጣምህ ‘you fasted’ is pronounced *tomh* or *toməh*, but ጣምህ ‘your fast’ is pronounced only *toməh*.

21.15.1.8. *Final ə*

No Amharic word ends in the vowel *ə* except the interrogative particle -ኝ -*nə*, which is pronounced with a rising intonation, and the interjection እህ

əhə used by a listener to encourage the speaker to continue talking. Likewise in poetry, for reasons of prosody, if the final word of a line ends in a consonant, a vowel ə may be added.

For a prothetic ə, see § 21.18.

21.16. Vocalic length

The Amharic vowels are short. Vocalic length occurs only occasionally. Thus, the combination of a final -a with the enclitic -a may result in *ā*: e.g., *käbbädä yännatun mot täräddä* (from *tärädda-a*) ‘so, finally Käbbädä learned of his mother’s death’.

The vowel is likewise long in *näyī*, the equivalent of *nāw əngī* ‘so it is!’; in a form such as *sämmū wāy* ‘you, there! do you hear me!’; and in *āy* ‘no’, as in *məsahən bällah? āy* ‘have you had lunch?’ (lit. ‘have you eaten your lunch?’). ‘No’.

A long vowel is also connected with a special tone in the occurrences of a verbal noun at the end of the sentence. Examples: *zənam limäṭa nāw, mən yəššalall? tolo tolo mähēd* ‘it is starting to rain. What is best (to do)? Walk fast’; *səfra lāmagnāt bāgize männäsāt* ‘to get a seat one should start out early’; *šāhafīwa səra fätta səttəččawwāt alāqawa kätāf malāt* (with long *ā*) ‘when the secretary was idly playing about, the boss suddenly showed up’.

21.17. Allophonic vowels

21.17.1. *ā*

The vowel *ā* has a phonetic variant *ā̃* (that is, a sound tending toward *o*) when preceded by a *w* or by a labialized consonant ending in *ä*, such as *q"ā̃*: thus, *wāndəmm* ‘brother’, *wāmbār* ‘chair’, *tāwāllädä* ‘he was born’, *q"ārrätä* as against *läbbäsä*.

This feature is also expressed in the orthography. Indeed, one finds numerous words spelled with the vowel *o* after *w*. Examples: *wäba* and *woba* ‘malaria mosquito’, *wāf* and *wof* ‘bird’, *wāmo* and *womo* ‘Omo River’, *wādoma* and *wodoma* ‘plant that causes a goiterlike swelling in the necks of sheep’.

21.17.2. *ə > ũ*

Likewise, the vowel *ə* has a variant *ũ* (that is, a sound tending toward *u*) when preceded by a *w*: e.g., *wūsäd* ‘take!’ as against *ləbäs*. The final *ə* of a labiovelar (such as *q"ə*) also has a variant *ũ*: e.g., *q"urät* ‘cut!’ as against *ləbäs*.

Some words are written either with the vowel *ə* or with the vowel *u*: thus, *wəma* and *wuma* ‘symbol’, *wəha* and *wuha* ‘water’.

21.18. Prothetic *ə* and loss of initial *ə*

21.18.1. Prothetic *ə* before *r*

A prothetic *ə* often occurs before initial *r*, rarely before the sibilants *s*, *š*, *z*, and occasionally before other consonants.

The prothetic *ə* precedes *r* mostly if the initial *r* has the vowel *ə*. Examples: *rəsas*, *ərsas* ‘lead, pencil’, *rəguz*, *ərguz* ‘pregnant’, *rəgo*, *ərgo* ‘coagulated milk’.

It also occurs with *r* having a vowel other than *ə*: thus, *rab*, *ərab* ‘hunger’, *ras*, *əras* ‘head’, *rat*, *ərat* ‘dinner’.

A prothetic *ə* without an alternate form occurs in *ərsu* ‘he’, *ərs^ua* ‘she’, *ərsaččäw* ‘they’.

21.18.2. Before other consonants

Examples for *s*: *sar*, *əsar* ‘grass’, *set*, *əset* ‘woman’.

Examples for *š*: *šəkokko*, *əškokko* ‘rock hyrax’, *šät*, *əšät* ‘almost mature grain’, *šok*, *əšok* ‘thorn’.

Before other consonants such as *t*, *l*, and *m*. Examples for *t*: *tub*, *ətub* ‘brick’, *tiy*, *əti* ‘gland’.

Example for *l*: *əlqəmoš* ‘game consisting of catching stones tossed up in the air’ (from *läqqämä* ‘collect’).

Examples for *m*: *mäbällät*, *əməbällät* ‘elderly widow’, *märr*, *əmmärr*, exclamation to urge one’s mount to jump over an obstacle.

21.18.3. Loss of initial *ə*

Loss of original *ə* (going back to *ʔ*, *ʕ*) occurs in *ʔat* ‘finger’ for *əʔat* (going back to Ge‘ez *ʔašbāʕət*, Semitic root *šbf*); *sat* ‘fire’ for *əsat*; *zab* ‘reins’ for *əzab*; *rat* ‘dinner’ for *ərat*; *məs* for *əms* ‘vagina’; *wənät* for *əwnät* ‘truth’; *qubat* ‘concubine’ for *əqubat*.

21.19. Meeting of vowels

As a rule, Amharic avoids the coming together of two vowels in pronunciation as well as in writing. If, for morphological reasons, two vowels should come together, the following may occur:

1. Elision of one of the vowels.
2. Introduction of a glide semivowel *w* or *y* between the two vowels.

21.19.1. Elision

Elision of one of the vowels occurs: (1) if the two vowels are the same; (2) if the 1st vowel is a central vowel and the 2nd vowel is other than a back vowel.

21.19.1.1. Same vowels

a-a > *a*: **sämma-aččäw* ‘he heard them’ > *sämmaččäw*; *asa aṭmağ* ‘fisherman’ > *asaṭmağ*.

ə-ə > *ə*: **lə-əsäbər* ‘so that I break’ > *ləsäbər*.

o-o > *o*: either elision of one *o* (**bäqlo-očč* ‘mules’ > *bäqločč*) or a semivowel *w* is produced (*bäqlo^wočč*).

21.19.1.2. Different vowels

If the 1st vowel is a central vowel *ä*, *a*, or *ə*, and the 2nd vowel is a vowel other than a back vowel *o* or *u*, one of the vowels is elided. Thus,

ä-ə > *ä*: **yä-əğğ* ‘of the hand’ > *yäğğ*

ä-a > *a*: **yä-amarəñña* ‘of Amharic’ > *yamarəñña*

a-ə > *a*: *amsa əgər* ‘centipede’ > *amsagər*

a-ä > *ä*: *ṭena-äñña* ‘healthy’ > *ṭenäñña*

ə-a > *a*: **sə-alf* ‘while I pass’ > *salf*; **yə-awqall* ‘he knows’ > *yawqall*;

**ə-alläbbät* becomes *alläbbät* ‘that which is in it’.

A form such as *šaläqa* ‘major’ arose from an original *ši-aläqa* becoming **šə-aläqa* > *šaläqa*.

21.19.1.3. With plural marker

In a plural noun, any final vowel of a singular noun may be elided when the plural marker *-očč* is added to it. Thus, **tāmari-očč* ‘students’ > *tāmaročč*; **wəšša-očč* ‘dogs’ > *wəššočč*. Note, however, that the vowel may be kept and a semivowel is then produced: thus, *tāmari^wočč*, *tāmari^yočč* (see below).

21.19.2. Insertion of a semivowel *w*, *y*

A semivowel *w* or *y* is inserted under the following conditions:

1. If the 1st vowel is a back vowel *u* or *o*, the semivowel *w* is inserted.
2. If the 1st vowel is a front vowel *i* or *e* and the 2nd vowel is the central vowel *a*, the semivowel *w* is inserted.

3. If the 1st vowel is a front vowel *i* or *e* and the 2nd vowel is the back vowel *o*, a semivowel *w* or *y* is inserted.

4. If the 1st vowel is a central vowel and the 2nd vowel is the back vowel *o*, a semivowel *w* is inserted.

Only a few occurrences will be given.

o-a > *o^wa*, *-wa* (without *o*): **bāqlo-aččən* > *bāqlo^waččən* or *bāql^waččən*

u-a > *u^wa* or *wa* (without *u*): **nāggāru-aččāw* ‘they told them’ > *nāggāru^waččāw* or *nāggār^waččāw*

i-o > *i^{vo}*, *i^{wo}*: **tāmari-očč* ‘students’ > *tāmari^{vo}očč*, *tāmari^{wo}očč*

e-o > *e^{vo}*, *e^{wo}*: **bāre-očč* ‘oxen’ > *bāre^{vo}očč*, *bāre^{wo}očč*

a-o > *a^{wo}*: **geta-očč* ‘masters’ > *geta^{wo}očč* (but also *getočč*)

21.20. Contraction of syllables

21.20.1. General

The succession *-Cāyā-* becomes *Ce*: thus, *wādet* ‘where to?’ from *wādā-yāt*; *āndet* ‘how?’ from *āndā-yāt*.

Cāyə- becomes *Ci*: e.g., *āndih* ‘such’ from *āndā-yəh*, lit. ‘like this’; *āndičč* ‘like this’ (fem.) from *āndā-yəčč* (*yəhəčč*).

The syllable *Cāy-* may be contracted in pronunciation and in spelling into *Ce*: e.g., *sāyf* and *sef* ‘sword’; *wāyzāro* and *wezāro* ‘madam, lady’.

Cyā becomes *Ce*: e.g., *gəddellāš* and *gəddyällāš* ‘carefree, unworried’.

Cyə becomes *Ci*: e.g., *asizo* ‘he having someone seized’ from *asyəzo*; (in proper names) *bārrihun* (from *bārr yəhun*) ‘may he be the defense’.

Cəyə becomes *Ci*: e.g., *sisābər* ‘when he breaks’ from *səyəsābər*; *bināgru* ‘if they tell’ from *bəyənāgru*.

Cāw becomes *Co*: e.g., *lāwz* and *loz* ‘almond’; *mābrād* ‘file’ (from Arabic) > **māwrād* > *morād*.

Cāwā may become *Co*: e.g., *dāmāwāz* and *dāmoz* ‘salary’ (lit. ‘blood and sweat’); *zāwātər* and *zotər* ‘always’.

Cwā may become *Co*: e.g., *təlwāt* and *təlot* ‘dowry’; *ləbb wāllād* and *ləb-bollād* ‘fiction’; *aswāssādā* and *asossādā* ‘have someone take’.

Cəw may become *Cu*: thus, *mulād* ‘place of origin’ (from *məwlad*); *həwkāt* and *hukāt* ‘disturbance’.

21.20.2. Final *Vya* > *a*

bāzziya and *bāzza* ‘there’; *lāzziya* and *lāzza* ‘for that’; *mankiya* and *manka* ‘spoon’; *ṭariya* and *ṭara* ‘roof’; *maldāya* and *malda* ‘brass arm bracelet’; *waləya* and *wala* ‘Walia ibex’; *frəmbiya* ‘cow’s chest’ and *frəmba*.

21.21. Alternation of vowels

21.21.1. Alternation *i-ə*

These two vowels are occasionally in free variation, particularly with the palatal *š* and occasionally with other palatals: thus, *šəro* and *širo* ‘stew made from the flour of parched beans’; *ğər* and *ğir* ‘a small military following a noblemen’; *māññəta* and *māññita* ‘sleep, slumber’.

The forms of the perfect 2nd fem. singular ending in *-š*, or 3rd fem. singular ending in *-äčč*, are combined with the object suffix pronouns through the vowel *i* or *ə*: thus, and *säddäbsəññ* and *säddäbsiññ* ‘you (fem.) insulted me’; *säddäbäččəññ* ‘she insulted me’ and *säddäbäččiññ*.

With other consonants: *dəda* and *dida* ‘dumb, mute’; *məst*, *məšt* and *mist* ‘wife’; *məyazəya* and *miyazəya* ‘May’.

The instrumental is also written and pronounced *mängäriya* and *mängärəya*.

21.21.2. Alternation *e-ä*

In the gerund 1st person singular, the regular final vowel *-e* (as in *nägərre*) alternates with *ä* in verbs ending in a palatal or in *y*. Thus, *käfəčče* and *käfəččä* ‘I opening’; *läbəšše* and *läbəššä* ‘I getting dressed’; *bəyye* and *bəyyä* ‘I saying’.

21.21.3. Initial *e, i, ə, o, u*

Some nouns with an initial vowel *e, i*, or *ə* have variants with initial *y* + vowel. Examples: *eli* ‘tortoise’ and *yäli*; *ečč*, expression of disgust, and *yäčč*; *itot* ‘nun’ and *yətot*; *imənt* ‘a nobody’ and *yəmənt*.

In Arabic loanwords: *yəməma* ‘turban’, from Ar. *šimāma*; *yəmam* (also *imam*) ‘leader of the Friday service’, from Ar. *šimām*; *yəmana* ‘money left in trust’, from Ar. *šimāna*; *yələma* ‘target’, and *ilama*, from Ar. *šilāma*.

A few nouns with initial *o* or *u* have variants with *w*: *ona* ‘bare place without any people’, and *wāna*; *of* ‘bird’, and *wäf*, *wof*; *uqabe* ‘guardian spirit’, and *wəqabe*.

21.22. Vowel harmony

21.22.1. Central vowels become rounded

When a rounded vowel *u* or *o* occurs in the root, the tendency is to harmonize the vowels *ə* and *ä* with the rounded vowels. Thus, the sequence *ə-u* may

become *u-u*: *qəmburs* and *qumburs* ‘fat white grub’; *buruk* ‘blessed’ instead of *bəruk*; *səlluse* and *sulluse* ‘ornamental collar for mules’.

ə-o may become *u-o*: *məšo* and *mušo* ‘dirge’; *šəro* and *šuro* ‘flour of roasted peas, sauce made from such flour’.

u-ə may become *u-u*: *šulləda* and *šulluda* ‘flesh of the thigh’; *šurrəbba* and *šurrubba* ‘braided hairdo’.

ä-o may become *u-o* or *o-o*: *mägogo* and *mugogo* (also *məgogo*) ‘griddle’; *tälo* and *tolo* ‘soon’.

o-ä may become *o-o*: *wärräta* (pronounced *worräta*) and *worrota* ‘benefit, favor’.

21.22.2. Rounded vowels become centralized

Note that, because for some of the above-mentioned examples the origin is unknown, it is quite possible that the rounded vowel *u* or *o* was the original one and it became dissimilated into *ə* owing to the preceding or following *u*: thus, an original *šulluda* may have become *šulləda*.

21.23. Syllabic structure

To illustrate the syllabic structure, the symbol C is used for a consonant, *C̣* for a geminated consonant, and V for a vowel.

The following syllables occur in Amharic:

V:	<i>u</i> , exclamation of surprise
VC:	<i>af</i> ‘mouth’
V <i>C̣</i> :	<i>əḡḡ</i> ‘hand’
VCC:	<i>ərf</i> ‘plow beam and handle of the plow’
CV:	<i>na</i> ‘come!’
CVC:	<i>mot</i> ‘death’
CV <i>C̣</i> :	<i>ləkk</i> ‘correct, size’
CVCC:	<i>səlt</i> ‘method’

These syllables may constitute words by themselves as in the examples cited above, or they may be part of multisyllable words.

An Amharic word may begin with any consonant, although initial *ɲ* is very rare. As for *p* and *ɸ*, they are rare and are usually of foreign origin.

An Amharic word may begin with any vowel except *ä*. The only exception to this rule is the word *ärä* ‘oh, really!’.

An Amharic word may end in any vowel. A final *ə* occurs only in the interrogative particle *nə* and in poetry for reasons of rhyme.

21.24. Consonant clusters

A consonant cluster is a succession of two consonants not separated by a vowel. Treatment of a consonant cluster is different in nouns and verbs.

21.24.1. Initial consonant clusters

Amharic has no initial consonant clusters. A word such as ንጉሥ is to be read *nəgus*; ክብር is *kəbur*. If, however, the 1st consonant is *k* or *g* and the 2nd consonant is *ɾ* and rarely *l*, or if the 1st consonant is *b* and the 2nd consonant is *l* or *ɾ*, or if the 1st consonant is *f* and the 2nd consonant is *ɾ*, there may or may not be a consonant cluster: thus, *krəstənnä* 'Christianity', *krämt* 'rainy season', *krar* 'six-stringed lyre'; *gräññ* 'left-handed'; *g(ə)rəmbiʔ* 'contrary (person)', *b(ə)latta*, honorific title, *b(ə)len* 'pupil of the eye', *b(ə)rälle* 'small carafe', *b(ə)rät* 'iron', *b(ə)rəndo* 'meat eaten raw', but *bəlat* 'stratagem', *bələh* 'clever'; *frida* 'young gelded steer fattened for slaughter', but *fəre* 'fruit'.

Note that an initial consonant cluster in a foreign word is avoided by a prothetic *ə*: thus *əsport* for *sport*; *əspil* 'pin' from Italian *spillo*.

21.24.2. Final clusters

Amharic has final clusters of two consonants in verbal forms regardless of the nature of the consonants.

Only a few principles will be discussed here.

In the verbs with identical 2nd and 3rd radical, there is clustering: e.g. ሊወድ *liwädd*.

If the 2nd radical is a liquid *l*, *ɾ*, *n*, there is clustering: *liwäld*, *bəlt*, *liwärd*, *lizänb*.

If the 2nd radical is a labial, there is clustering: *linäfs*, *lilämd*, but if the 3rd radical is a liquid, there may or may not be clustering: *lisäbr* or *lisäbər*.

If the 2nd radical is a liquid and the 3rd radical is a labial, there is clustering: *likärm*.

If the 2d radical is a velar and the 3d radical is a labial, there is hesitation: *yərägm* and *yərägəm*.

If the last radical is a sibilant, there may or may not be clustering: *liläbs* or *liläbəs*.

If the last radical is a dental, there is clustering: *yəbält*, *käbt*, *fərd*.

One would have to investigate whether there is a difference between nouns and verbs.

A geminated consonant is considered double. A word such as ልቅዎ, with geminated *q*, cannot, therefore, be read *ləqqm*, with the final cluster *-qqm*. It has to be read *ləqqəm*.

There is no final cluster of three consonants: thus መንግሥት is to be pronounced *mängəst* and not *mängst*.

21.24.3. Medial clusters

When two consonants meet in the middle of a word, the 1st consonant closes the syllable and the 2nd consonant opens the next syllable. Thus, ይንገር reads *yən-gär*; መንግሥት *män-gəst*.

The excluded syllabic pattern mentioned above is helpful in reading words that have consonants in the 6th order.⁹ Thus, for instance, ድንግል can be read only *dəngəl* and not *dənəgəl*, nor can it be read *dəngl* as no final cluster of three consonants exists (see above). A word such as መጥፎ can be read only *mätfo* and not *mätəfo*. A word such as ቡርቱ can be read only *bərtu* and not *bərətu*.

This clustering in medial position is valid only for nouns; for verbs the clustering depends on the verb form. Thus, ይስማል 'he kisses' is to be read *yəsəmall* and not *yəsmall*; or አድርጎ 'he having done' is to be read either *adərgo* or *adrəgo*.

There are no clusters of three consonants in medial position. The word መንግሥት cannot, therefore, be read *mängsət*. For its reading *mängəst*, see above.

A consonant in the 6th order preceding a geminated consonant is to be followed by *ə*. Thus, ምልክት (where *k* is geminated) is to be read *mələkkət*. Indeed, a reading such as *məlkkət* would result in a medial cluster of three consonants.

There are no definite rules covering all the possibilities. Thus, for instance, while ልክስክስ is to be read *ləkəskəs*, there is no valid rule that it should not be read *ləksəks*.

21.24.4. Clustering in phrases

All the examples of syllabic structure have dealt with words in isolation. Modifications, however, occur when words constitute a phrase. For instance, ሌብስ is pronounced *ləbs* in isolation, the last consonant having no vowel. In the phrase ሌብስ፡ኑው *ləbs nəw*, however, a vowel *ə* is intercalated

9. The consonant in the so-called 6th order [vowel differentiation of the consonant letter] either has the vowel *ə* or is *zero*.

between the last consonant of *ləbs* and the initial *n* of *nāw*, and the phrase is then read *ləbsə nāw*; ነቅ፡ነው *näččə nāw* ‘it is white’; አንድ፡ብር *andə bərr* ‘one dollar’. Indeed, these two words form a unit that would then have a medial cluster of three consonants, a clustering not found in Amharic.

21.25. Accent

21.25.1. Stress

In general Amharic has an almost even distribution of stress on each syllable. Bold type indicates a stressed syllable.

It is safe to state that the last syllable is not stressed: thus, *səga* ‘flesh’ (and not *səga*); *wāmbār* ‘chair’ (and not *wāmbār*).

In bisyllabic nouns the stress is on the 1st syllable: thus *sāñño* ‘Monday’, *sānde* ‘wheat’.

In trisyllabic nouns the stress seems to vary: thus, *samuna* ‘soap’, but *tārākāz*, *zəng^wərg^wər* ‘variegated’.

In quadrisyllabic nouns the stress is on the penultimate syllable: thus, *zānāzāna* ‘pestle’, *arāng^wade* ‘green’.

The syllable preceding a geminated syllable is likely to be stressed: thus, *fällägä*, *fällallägä*, *yəfälləgal*; *mäsäkkärä*, *yəməsäkkər*.

The question of the accent in Amharic still awaits a thorough investigation.

21.25.2. Intonation

Intonation plays an important role in the language. A few examples will illustrate its occurrence. Thus, *lə* + imperfect with a rising tone may express surprise: e.g., *ləttəhed?* ‘are you really going to leave?’

The gerund at the end of a sentence may be uttered with rising-falling tone: e.g., *käbbädä yät allä? hedo* ‘where is Käbbädä? Why, he has already left’.

mən ‘what?’ or *yāmən* ‘of what?’ followed by a noun has a wide variety of meanings depending on the intonation: e.g., *zare dägmo mən* (or *yāmən*) *sāw nāw?* ‘what kind of strange (unwanted, unexpected) person came today?’; *yəh yāmən səra nāw?* ‘what sort of job is this?’ (with a shade of contempt or mock amazement).

The *tä*-stem may have an overtone of irony or sarcasm. The utterance is then spoken with a rising intonation. Thus, *suf täläbbäsänna yəkk^wärall* ‘(look at him,) he puts on a wool suit and is all vain’.

Bibliography

- Afework, G. J. 1905. *Grammatica della lingua amarica*. Rome: Tipografia della R. Accademia dei Lincei.
- Armbruster, C. H. 1908. *Initia amharica: An introduction to spoken Amharic*. Part I, *Grammar*. Cambridge: Cambridge University Press.
- Cohen, Marcel. 1936. *Traité de langue amharique* (Travaux et mémoires de l'Institut d'ethnologie 24). Paris. Institut d'ethnologie.
- . 1939. *Nouvelles études d'éthiopien méridional*. Paris: Champion. [Amharic, pp. 1–371.]
- Dawkins, C. H. 1969. *The Fundamentals of Amharic*, rev. ed. Addis Ababa: Sudan Interior Mission. (1st ed., 1960).
- Guidi, Ignazio. 1889. *Grammatica elementare della lingua amarica*. Rome: Tipografia della R. Accademia dei Lincei.
- Hartmann, J. 1980. *Amharische Grammatik*. (Aethiopistische Forschungen 3). Wiesbaden: Steiner.
- Leslau, Wolf. 1995. *Reference Grammar of Amharic*. Wiesbaden: Harrassowitz.
- Ludolf, Hiob. 1698. *Grammatica linguae amharicae*. Frankfurt am Main: Prostat apud Johannem David Zunnerum.
- Podolsky, Baruch. 1991. *Historical Phonetics of Amharic*. Tel Aviv: author.
- Praetorius, Franz. 1879. *Die amharische Sprache*. Halle: Waisenhaus.

Egyptian and Coptic Phonology

Antonio Loprieno

University of California, Los Angeles

22.1. Introduction

At the present state of our knowledge, a discussion of Egyptian and Coptic phonology must be addressed primarily as an issue of *diachronic*, rather than synchronic linguistics. While it is possible to recognize regular patterns of sound change in the history of the Egyptian language as a whole, including in many cases its Afroasiatic antecedents, the synchronic systems of phonological oppositions at any given time in the four millennia of the productive history of this language often defy a clear analysis. Furthermore, the dynamic models of historical phonology tend to hide many uncertainties behind the regularity of a reconstructed paradigm, conveying the misleading impression that for each of the different phases of the language (Early Egyptian, Middle Egyptian, Late Egyptian, Coptic) we are indeed able to establish a discrete phonological system.

The phonetic realities underneath the abstract phonological reconstructions are even more elusive: the traditional pronunciation and transliteration of many hieroglyphic phonemes rest upon hardly anything more than scholarly conventions, and even for the relatively well-known Coptic, in which Egyptian sounds are rendered in a Greek-based alphabet, it is difficult to assess reliable phonetic values for some of the Greek signs and of the Demotic graphemes that were added to the Greek alphabetic set.

In fact, the main reason for the difficulties in reconstructing the phonology of Ancient Egyptian lies in the very nature of the writing systems: Hieroglyphs, Hieratic, and Demotic represent the mere consonantal skeleton of a word (and sometimes only a portion thereof), followed by indicators of lexical classes, the so-called “determinatives.” Semivocalic phonemes are rarely indicated, vowels practically never. As for Coptic, in which vowels are indeed rendered, one should not underestimate the methodological difficulty inherent in the widespread assumption of a phonological or phonetic identity between a specific Coptic sign and its original value in the Greek system—an identity which is by no means unquestionable.

Therefore, the reconstruction of the phonological inventory and of the phonetic values in any period of the history of Egyptian is bound to remain highly hypothetical: it can only be approached through a heuristic procedure in which three dimensions are checked against each other and mutually verified (cf. § 22.2): the reconstruction of Afroasiatic prehistory,¹ the information drawn from contemporary sources in other (mostly Semitic) languages with a better investigated phonology (Hoch 1991), and the laws of phonological evolution leading from older Egyptian to Coptic.²

22.2. Heuristic criteria

In spite of these difficulties, the study of Egyptian phonology has achieved significant progress since the initial studies of the late 19th century, both in the assessment of consonantal values and in the reconstruction of vocalic phonemes and prosodic rules. To achieve this goal, scholars rely on four procedures of linguistic reconstruction.³

22.2.1. *Comparative Afroasiatic linguistics*

Egyptian is a language of the Afroasiatic phylum, and the presence of established etymological correspondences offers a fundamental source for the reconstruction of phonological values. E.g., since Eg. <q3b> corresponds to Sem. *qrb* meaning ‘interior part’, one can confidently establish that Eg. <q> = /q/ and that = /b/.

22.2.2. *Contemporary transcriptions in foreign languages*

Many Akkadian texts, especially from the archive of el-‘Amarna (15th–14th c. B.C.E.), contain Egyptian words, names, and short phrases in cuneiform transcription. Although the phonology and the graphemics of Akkadian are themselves by no means fully decoded, these transcriptions provide a valuable insight into the contemporary pronunciation of Egyptian. E.g., Eg. <stpnr’> ‘the-one-whom-(the-god-)Ra-has-chosen’ (royal name of King Ramses II) appears in cuneiform as *šá-te-ep-na-ri/e-a*, a form on the basis of

1. Suggestions for the reconstruction of the phonological evolution from Afroasiatic to Egyptian are presented by Schenkel 1990: 48–57; Kammerzell 1992; and Zeidler 1992.

2. The most complete description of these rules and of the patterns of Egyptian vocalization is found in Osing 1976: 10–30.

3. Cf. Schenkel 1990: 23–28. This book presents the most up-to-date and compact picture of Egyptian phonology (pp. 24–93). I shall make specific references to it only in the rare cases in which my analysis differs from Schenkel’s in a significant way.

which one can both posit the contemporary Egyptian pronunciation as */sa,tepnə'ri:ʕa/ and observe the correspondence Eg. <s> // Akk. <š>, both of which were probably realized as [s] or as a sound very close to it (at least in some dialects).⁴

22.2.3. Egyptian renderings of foreign words, especially of Northwest Semitic origin

This criterion represents the symmetrical counterpart to the preceding one: it provides an insight into the phonology of contemporary Egyptian while at the same time offering the possibility of verifying scholarly assumptions on Semitic phonology. E.g., Northwest Sem. **sōpēr* 'scribe' ⇒ Eg. <tu-pa-r>. The relevance of this piece of evidence is twofold: on the one hand, it raises questions about the phonological status and the phonetic realization of Eg. /c/, which is the palatal phoneme usually transcribed *t* by Egyptologists, while on the other, it can also be used to shed some light on the value of the phoneme /s/ (*samekh*), which originally must have been an affricate [tʃ] in Semitic (cf. Faber 1990: 627; Hoch 1991: 484f.; Faber 1992).

22.2.4. The evidence provided by Coptic

The latest stage in the development of Egyptian provides the broadest basis for the study of the phonology of older periods of the language as well. E.g., Eg. <w'b> 'pure', 'to be pure', 'priest' appears in Coptic in the lexemes Ⲡⲩⲁⲁⲃ 'holy', Ⲡⲩⲟⲡ 'to be pure', Ⲡⲩⲏⲏⲃ 'priest'. This evidence enables us to reconstruct three different vocalization patterns underlying the same graphic reality of hieroglyphic Egyptian: the stative **wā'ḥaw* 'he is pure', the infinitive **wa'āb* 'to become pure', and the noun **wī'ab* 'priest'. At the same time, this piece of evidence raises questions of consonantism, i.e., the fate of the phoneme /ʕ/ and the reason for the alternance ⲃ vs. ⲡ in the Coptic forms as opposed to in both cases in their Egyptian antecedents.

In the practice of Egyptian phonological reconstruction, these four aspects appear constantly combined: while each of them, if considered individually, proves largely inadequate in order to determine a synchronic stage, together they convey a relatively homogeneous picture of the fundamental laws of Egyptian phonological development. What follows in the next paragraphs (§§ 22.3–6) is a tentative historical phonology of Egyptian from its Afroasiatic roots to alphabetic Coptic. Transcriptions from Egyptian and

4. See Faber 1990: 627ff.; 1992. For dialectal differences in the case of Akk. *š* cf. von Soden 1969 § 30.

Semitic follow the established conventions in these respective fields and are rendered in *italics*; transliterations of graphemes without reference to their phonological status are indicated in angle brackets (<x>); phonemes (/x/) and tentative phonetic values ([x]) are represented according to IPA conventions.

At this point, a methodological warning is in order: in the case of Egyptian (and of many other ‘philological’ languages known only through written records), the distinction between the *phoneme* as the distinctive minimal unit of the language (/x/) and the often much larger inventory of *sounds* ([x]) representing its physical realizations is heuristically less practicable than for languages with a better-known phonological structure: while scholars can strive for the reconstruction of the “sound units” of the language, the technical assessment of their phonological status, which would require in each case the minimal pair test, often proves a very problematic endeavor: on the one hand, our only source of information is represented by a complex writing system which combines phonetic and semantic principles; on the other hand, because of the restrictiveness of cultural conventions governing the use of writing in Egyptian society,⁵ our knowledge of the lexicon is doomed to remain far from exhaustive.

22.3. The prehistory of Egyptian phonology

Before the emergence of Egyptian as a written language, a few adjustments within the stock of phonemes inherited from “Afroasiatic”⁶ seem to have taken place. Three major evolutions from the original phonological stock characterize the Egyptian domain as it begins its recorded history.⁷

(a) In the apical and interdental series, voiced *d, *z, and *ð develop into the pharyngeal phoneme /ʕ/ (Rössler 1971: 275–77), probably going through an intermediate stage with pharyngealized lateral: *d, *z, *ð (> *ɫ) > /ʕ/.⁸ E.g., Eg. *ʕ.t* ‘portal’, cf. Sem. **dalt* ‘door’; Eg. *ʕ* ‘to speak a foreign

5. What is often referred to as “rules of decorum”: cf. Eyre and Baines 1989.

6. “Afroasiatic” is here used as a conventional term to indicate the set of linguistic features which Egyptian shares with a certain number of other language families (Semitic, Berber, Cushitic, Chadic), without implying the belief in the existence of an actual proto-language ancestral to these families. The different theoretical models are discussed in Loprieno 1986: 1–12, 187–90.

7. In the following transcriptions, *ǃ* denotes an unidentified short vowel (*a*, *i*, or *u*), corresponding (for typographic reasons) to the Egyptological convention *ˁ*.

8. Cf. the comparable evolution from Proto-Sem. **d* to Aram. <q>, later <ʕ>: **ʕd* > <ʕarqā> > <ʕarā> ‘earth’ (Brockelmann 1908: 134).

language', cf. Sem. **lǧz* (Ar. *laǧaza* 'to speak enigmatically', Hebr. *l'z* 'to speak a foreign language'); AA **ḏup̄p̄* 'fly' > Eg. *ʔfj̄* */ʕuffv̄j/ > Coptic **Δϣ**, cf. Sem. **qbb* (Akk. *dubbum*, Ar. *qubāb*, Hebr. *zəbûb*).

(b) Among the liquids, the original opposition between nasal **n*, lateral **l*, and vibrant **r* underwent a profound reorganization, not yet fully understood in its specific details, in which a role was also played by dialectal varieties. AA **n* and **r* were kept as Eg. /n/ and /r/—the latter being the phoneme conventionally transcribed *ṣ* by Egyptologists and traditionally taken to be a variety of glottal stop /ʔ/, but in early Egyptian probably a "uvular trill";⁹ Eg. *jnk* */ja'nak/ (Kammerzell 1991b: 201), Sem. **ʔanāku*, 1st sg. pronoun, or Eg. *kəm* */karmv̄w/ (Osing 1976: 857), Sem. **karm* 'vineyard'. On the contrary, AA **l* does not display uniform Egyptian correspondences nor is Eg. */l/ indicated by an independent grapheme, in spite of its almost certain presence in the phonological inventory of the language: AA **l* corresponds to Eg. <n>, e.g., AA **lis* 'tongue' > Eg. *ns* */lis/, cf. Coptic **ΛΔϢ**, Sem. **liš-ān*; to Eg. <r>, e.g., *jzr* */jaθr̄v̄w/ 'tamarisk', cf. Sem. **ʔat̄l̄*; to Eg. <ɜ>, e.g., *ɜ* 'to speak foreign languages', cf. Sem. **lǧz* (see above); and to Eg. <j>, e.g., AA **lib* 'heart' > Eg. *jb* */jib/, cf. Sem. **libb* or AA **lwn* 'color' > Eg. *jwn* */ja'win/ (Osing 1976: 316), cf. Sem. **lawn*. Presumably, proto-Eg. **l* merged with other sonorants in the dialect which eventually led to the written language, while still being kept in less normative varieties of the language: in the New Kingdom, when Later Egyptian became the written form of the language for the domain of administration and literature, a specific grapheme <n>+<r> was created in order to express the phoneme /l/. In Demotic, /l/ is autonomously indicated by a grapheme <l>, a diacritic variety of <r> = /r/.

(c) The AA velar plosives **k*, **g*, and **q* display two outcomes in Eg.: either they are maintained as *k* /k/, *g* /g/, and *q* /q/, or they are palatalized into *t* /c/, *j* /j/, and *ḏ* /j/ respectively: cf. the 2nd person suffix pronoun masc. /k/ < **-ka/-ku* vs. fem. /c/ < **-ki* (Kammerzell 1991b: 198ff.) or the opposition between the two Eg. roots *wɜḏ* (cf. *wɜḏ* */wa:riḏ/ 'green') and *jɜq* (cf. *jɜq.t* */jurqat/ 'vegetables') from an identical AA root **wrk*.

(d) The phonemes corresponding to the so-called "emphatic" series of other branches of the AA phylum lost their phonological status in Egyptian, merging either with the corresponding voiceless fricative, as in the labial

9. A possible remnant of the early pronunciation of this phoneme is perhaps its outcome as Coptic /r/ in specific phonetic surroundings: ^Bⲭⲣⲟⲃⲓ 'sickle' < *h b.t* */ḥa'rabjv̄t/(ʔ), with [ḥr] > [kʔr]. Cf. the references in Westendorf 1965: 67.

series, in which AA **p* develops into Eg. /f/: AA **sp̄y* ‘seven’ > Eg. *sfḥw* */safḥaw/, cf. Sem. **šb*; or with the corresponding voiced plosive: (1) the AA emphatic dentals **t* and **s* merge into Eg. /d/: Eg. *dwn* ‘to stretch’ */da:wan/, cf. Sem. *ṭwl* ‘to be long’; Eg. *wdpw* ‘servant’, cf. Ar. *waṣīf*; (2) in specific phonetic environments, the AA emphatic velars **k* and **x* merge into the voiced palatal stop /j/, the phoneme conventionally transcribed *ḏ* by Egyptologists: AA **wrk* > Eg. *wṣḏ* */wa:riḏ/ ‘green’, cf. Sem. **warq* ‘leaf’; AA **nxm* > Eg. *nḏm* */na:jiṃ/ ‘sweet’, cf. Sem. **n’m*. As we just saw, in absence of palatalization, AA **k* is kept in Eg. as /q/, which was probably articulated as ejective [qʰ] (see § 22.6 for Coptic evidence of this articulation): from AA **kṛb/kḭb* are derived both Eg. *qṣb* ‘interior’ (cf. Akk. *qerbum* ‘inside’) and Eg. *ḏnb* ‘to turn’ (cf. Ar. *qḭb* ‘to turn around’). As for AA **x*, when not subject to palatalization it merges into the voiceless pharyngeal fricative /ħ/: AA **ḫal* > Eg. *ḥr* */ħar/ ‘on’, cf. Sem. **al*.

22.4. The phonological system of Early Egyptian (about 2500 B.C.E.)

At the beginning of its written history, i.e., during the historical period known as the “Old Kingdom” (2800–2150 B.C.E.), one can assume that the Egyptian language displayed the phonological inventory shown in Table 22-1. Here, *x* indicates the traditional Egyptological transcription, /*x*/ the posited phoneme, [*x*] a tentative phonetic reconstruction (if different from /*x*/).

Some contemporary scholars, following Rössler 1971 (among Egyptologists cf. primarily Schenkel 1990: 24–57; cf. also Kammerzell 1992: 169ff.; Zeidler 1992: 204ff.) and a long tradition going back to the 19th century, offer a partially different analysis of these phonemes: since Eg. <*d*> and <*ḏ*> represent, as we just saw, the heirs of AA “emphatics” (**t/s* and **k/x* respectively), these phonemes, rather than as “voiced” /*d*/ and /*j*/, should be understood as “voiceless emphatic” <*d*> = /*t̥*/ and <*ḏ*> = /*ç*/, although the actual phonetic realization of the feature [+EMPHATIC], whether pharyngealization, velarization, or glottalization, cannot be determined with certainty (Kammerzell 1992: 169).

Yet, because of the presence of just *two*, rather than *three* phonemes in the respective Egyptian consonantal series, I prefer to analyze them as poles of the simpler binary opposition “voiceless” vs. “voiced”.¹⁰ However, an important discovery of the alternative approach to Egyptian phonology must

10. An excellent discussion of adequacy and advantages of this simpler solution is offered by Hoch 1991: 508 ff.

Table 22-1. Early Egyptian Consonants

	Bilabial	Labio-dental	Inter-dental	Dental	Alveo-palatal	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive										
Voiceless	<i>p</i> /p/ [p ^(b)]			<i>t</i> /t/ [t ^(b)]		<i>t̥</i> /c/ [c ^(b)]	<i>k</i> /k/ [k ^(b)]	<i>q</i> /q/ [q []]		<i>ʔ</i> ^a
Voiced	<i>b</i> /b/			<i>d</i> /d/ [t []]		<i>d̥</i> /j/ [c []]	<i>g</i> /g/ [k []]			
Fricative										
Voiceless		<i>f</i> /f/	<i>z</i> /θ/	<i>s</i> /s/ [s]	<i>š</i> /ʃ/	<i>h̥</i> /ç/		<i>ħ</i> /ħ/	<i>ʕ</i> /ʕ/	<i>h</i> /h/
Voiced										
Nasal	<i>m</i> /m/			<i>n</i> /n/						
Vibrant				<i>r</i> /r/ [r]				<i>ʕ</i> /ʕ/		
Lateral				<i>l</i> ^c						
Glide	<i>w</i> /w/					<i>j</i> /j/ ^d				

^aThe glottal stop [ʔ] was probably not a phoneme in Egyptian in its earliest phase; later on, presumably during the Middle Kingdom (2000–1750 B.C.E.), /ʔ/ represents on the one hand the result of the evolution *kl* > /ʔ/ (cf. note b), on the other hand the outcome of *ʃl* > /ʔ/ between two vowels in posttonic position (**ba:jn/* > **ba:ʔen/* ‘bad’) and before an unstressed vowel in initial position (**ʃa:nak/* > **ʔa:nak/* ‘f’). Kammerzell (1991b: 186–87; 1992: 168–69) prefers a consistent interpretation of *ʃ* > as palatal glide rather than as glottal stop /ʔ/.

^bIn the later phases of Early Egyptian (i.e., probably during the Middle Kingdom), the uvular trill *k*/, which is the Eg. heir of AA **r*, progressively tends to acquire the realization as glottal stop [ʔ]—an evolution which appears almost completed in the New Kingdom (1550–1050 B.C.E.); cf., however, n. 9.

^cIn the hieroglyphic system, the phoneme *l*/ is not indicated unambiguously: it is frequently conveyed by <ṯ> and <ṯ>, more rarely by < > and <ṯ>, cf. above.

^dFor the writing of this phoneme, the following general rules apply (with many exceptions): *ʃ* is rendered by <ṯ> in initial position: <ṯ> = **ʃa:tvj/* ‘father’, and immediately following a stressed vowel: <ṯjn> = **ba:jn/* ‘bad’; by <ṯj> within a word, if *ʃ* immediately precedes the stressed vowel: <ṯjkk> = **ʃaʕʕjak/* ‘you will appear’; by <ṯ> at the end of a word: <ṯt> = **ʃa:tvj/* ‘father’.

be borne in mind and accounted for: on the basis of both comparative evidence¹¹ and diachronic signals,¹² Egyptian *mediae* often appear to have neutralized the feature [+VOICED] and to have been realized—together with the uvular plosive /q/—as ejective stops.¹³ Ejectivity, the existence of which can also be inferred through indirect Coptic evidence (cf. § 22.6), brought these phonemes into the phonetic proximity of Semitic (and Afroasiatic) “emphatics”: most likely /d/ = [tʰ], probably also /j/ = [cʰ], /g/ = [kʰ], and /q/ = /qʰ/. A possible explanation of this phenomenon of especially initial devoicing¹⁴ is that the feature [+VOICED] must have progressively become redundant under the competition of the optional aspiration which, at least in some varieties of the language and specific environments, characterized Egyptian voiceless stops: /p/ = [pʰ] and /t/ = [tʰ], probably also /c/ = [cʰ] and /k/ = [kʰ].¹⁵ This is shown by the fact that Eg. /p/ and /t/ are rendered in the Greek transcriptions by φ and θ respectively: *pth* */pi'taħ/ ‘(the god) Ptah’ > Φθα, and Eg. /c/ and /k/ often by σ and χ respectively: *tb-ntr* */cab'na:car/ > */cəb'nu:tə/ ‘(the city of) Sebennytoś’ > Σεβεννυτος, *b3k-n-rn=f* */ba:rak-ʎn-ri:nvʃ/ > */bokko'ri:(nʃ)/ ‘Bocchoris’ (p.n., lit. ‘servant-of-his-name’) > Βογχορις,

11. Schenkel 1990: 33–41. In loanwords from Egyptian to Semitic, Eg. *d* is always rendered by Sem. *t*: Eg. *jdmj* */jv̥du:m̥j/ [jv̥t'u:m̥(ʃj)] > Hebr. *ʔēṭān* ‘red linen’. The same holds true for Babylonian transcriptions of Eg. words: *jfdw* */jaʔdaw/ [jəʔtʰaw] ‘four’ = Middle Bab. *iptau* (Lambdin 1952: 136–37); Sem. *t*, on the other hand, is rendered both by Eg. *d* (with which it shared “markedness,” whatever the phonetic realization of this feature may have been) and by Eg. *t* (with which it shared “voicelessness”). Also, /g/ and /q/ were articulated in a very similar way, i.e., [kʰ] and [qʰ] respectively, a fact which explains why Eg. *g* (= [kʰ]) is always rendered by Sem. *q* (= [q]): Eg. *gstj* */gastv̥j/ [ʔkʰast(v̥j)] ‘palette’ > Hebr. *qešet* (< **qašt*) ‘bow’ (Lambdin 1952: 148), whereas both Sem. *q* (= [q]) and *g* (= [g]) can be rendered by Eg. *g*. As for Eg. *ḏ* (“ejective”), it regularly corresponds to Sem. *š* (“emphatic”): *ḏn.t* */ʃuʃn̥v̥t/ [ʔcʰuʃn̥(v̥t)] ‘(the city of) Tanis’ > Hebr. *šōʾan* (< **šuʃn*). Cf. Hoch 1991: 512 ff.

12. Cf. the consistency displayed by the evolutions Eg. /d/ > Coptic τ, Eg. /j/ > Coptic x, Eg. /g/ > Coptic κ or ϣ: see Worrell 1934: 17–30.

13. For the discussion of similar “glottalic” approaches to the phonology of Indo-European and of the proximity of voiced phonemes to ejectives see Schmalstieg 1990: 362–65. An exception is represented by /b/ = [b], in which the feature [+VOICED] was probably kept because of the difficulty of maintaining in a linguistic system a glottalized [pʰ], due to the distance between glottis and lips: cf. the discussion by Schmalstieg (pp. 363 f.).

14. This pattern of devoicing represents a form of “initial strengthening” (Hock 1991: 162–64).

15. An excellent analysis of the relation between three different types of stops (voiced-unaspirated, voiceless-aspirated, and voiceless-unaspirated) is provided by Worrell 1934: 17 ff.: while Egyptian “voiceless” plosives are aspirated, their “voiced” counterparts, which were probably articulated as ejectives, correspond rather to Worrell’s “half-voiced” (i.e., voiceless-unaspirated) stops.

Βοκχορις, Βοχορινις. This aspiration is still exhibited by some Coptic dialects such as Bohairic (cf. § 22.6).

As for the sibilants, Old Kingdom Egyptian displays three phonemes, conventionally transcribed *z* (or *s*), *s* (or *š*), and *š*. When subject to palatalization, this last phoneme corresponds etymologically to AA **x* (which, as a rule, evolves to Eg. *ḥ* = /ç/): Eg. *ḥmm*, *šmm* ‘to become hot’, cf. Sem. **ḥmm*. This phenomenon seems indeed to indicate an articulation /ʃ/ for Eg. *š*, although both AA **š* and **ś* are continued by Eg. *s* (š), i.e., by the second phoneme listed above: cf. AA **šu*: ‘he’ > Eg. *sw* */suw/ (cf. Kammerzell 1991b: 190ff.), Sem. **šuwa*; AA **šapat* ‘lip’ > Eg. *sp.t* */sa:pat/ (Osing 1976: 870f.), Sem. **šapat*. It is quite possible, therefore, that Eg. *s* /s/ was characterized by a supplementary articulatory feature, whose precise phonic nature (perhaps of lateral or palatal type) is impossible to determine. Eg. *z*, on the other hand, is the heir of AA **θ* and **s*, as is shown by such correspondences as the already mentioned Eg. *jzr* */jaθr̥w/ ‘tamarisk’, cf. Sem. **ʔatl* or AA **sulḫam* ‘locust’ > Eg. *znḥmw* */θanḥu:m̥w/ (Osing 1976: 454), cf. Hebr. *sol’ām*. Here I reconstruct *z* as /θ/,¹⁶ but it needs to be stressed that the phonological opposition between /θ/ and /s/ was neutralized by the beginning of the Middle Kingdom, at which time <*z*> and <*s*> had become graphic variants of the same phoneme /s/. However, the articulation and the phonological status of sibilants in the whole phylum remains a thorny issue of AA linguistics.

The Eg. phoneme /j/ represents the outcome of AA **j* (Eg. *jmn* ‘right side’, therefore ‘West’, the point of reference being represented by the sources of the Nile, i.e., the South, vs. Sem. **ymn* ‘right side’, therefore ‘South’, the point of reference being the place where the sun rises, i.e., the East) and of AA **l* (Eg. *jwn* ‘color’, cf. Sem. **lawn*, see also § 22.3) when subject to palatalization. Probably by the beginning of the Middle Kingdom, as part of the global reorganization of liquid phonemes which took place in Egyptian (with /*r*/ > /ʔ/ and the neutralization of the opposition between /*l*/ and other sonorants, cf. Table 22-1), /*j*/ turned into /ʔ/ before an unstressed vowel in

16. Schenkel 1986 suggests the interpretation of *z* as affricate [t͡s], because it stands for /*t*/ + /*s*/ in the word *nzw* ‘king’, whose more traditional writing is *ntsw*. Whether an affricate (as suggested by Schenkel and by the equation with AA **s*) or a fricative (as suggested by the correspondence with AA **θ*), it is not surprising that this phoneme should be used to indicate a sibilant immediately following a nasal, a phonetic surrounding which often tends to generate affrication: /*ns*/ < <*nts*>, <*nz*> = [ˈn̥t͡s] (à la Schenkel) or <*nz*> = [ˈn̥θ] > <*nts*> = /*ns*/ [ˈn̥t͡s] (as defended here): for “consonantal epenthesis” (as in the case of [ns] > [n̥t͡s]) cf. Hock 1991: 117 ff.

initial position (e.g., *jwn* */ja'win/ > */ʔa'win/ 'color') and in postvocalic position following the stress (e.g., *ḥjpw* */ḥujpǝw/ > */ḥeʔpǝw/ '[the god] Apis').

Among the guttural fricatives, <ḥ> = /ç/ continues AA *x (AA **xanam* > Eg. *ḥnmw* '[the ram-god] Khnum', Ar. *ḡanam* 'sheep'), <ḥ> = /χ/ is the outcome of AA *ɣ (AA **wsy* 'wide' > Eg. *wsh*, Ar. *ws*), and <h> = /h/ derives from AA *ḫ when not subject to palatalization (AA **sulḫam* 'locust' > Eg. *znḥmw*, Hebr. *sol'ām*). The phoneme <h> = /h/ does not display any unequivocal AA cognate.

22.4.1. Vowels

The vocalic system of early Egyptian can be reconstructed as in Table 22-2.

Table 22-2. Early Egyptian Vowels

	Short	Long
Front	/i/	/i:/
Central	/a/	/a:/
Back	/u/	/u:/

The three vocalic qualities posited for early Egyptian are inherited directly from the AA prehistory of the language. While never spelled out in writing, vocalic phonemes can be reconstructed with a sufficient degree of systematic reliability on the basis of the four criteria formulated in § 22.2. For the earliest phase in the development of the Egyptian phonological system we do not posit the existence of a phoneme shwa.

Unlike stressed vocalic phonemes, unstressed vowels cannot be reconstructed with any degree of reliability. E.g., in *ntr* */na:car/ 'god', while the stressed vowel can be reconstructed directly from Coptic **ⲛⲟⲩⲧⲉ** (with */na:/ > /nu:/, cf. § 22.6), the quality of the unstressed vowel in *-car can only be inferred indirectly through the feminine form *ntr.t* */na'ca:rat/ > Coptic **ⲛⲧⲣⲉ** (with */ca:/ > /to:/, cf. § 22.6). The extent to which a whole paradigmatic class should be posited on the basis of analogy is still a matter of intense scholarly debate.

22.4.2. Syllabic structures

As a general rule, the opposition between short and long vowel is not phonological, but determined by the respective syllabic structure: long vowels appear in open stressed syllables, short vowels in closed and in open

unstressed syllables. Major exceptions are represented by the presence of a long vowel in closed stressed syllable in the infinitive of biconsonantal verbal roots and, at least according to some scholars, the possibility of long or doubly-closed syllables in final position. Accordingly, the following five or seven patterns of syllabic distribution (patterns 6–7 are not universally accepted) are characteristic for early Egyptian (C = consonant, V = vowel, # = word boundaries):

1. 'CVC *jnn* */ja'nan/ 'we'
2. ()CVC() *rmṯ* */ra:mac/ 'man'
3. 'CV: *ḥtp* */ḥa:tip/ 'pleasing'
4. #CV' *tpj* */ta'pij/ 'first'
5. #CV:C# *mn* */ma:n/ 'to stay'
6. _'CVCC#, _'CV:C# *mdw.w* */ma'duww/ 'words'¹⁷
7. _'CV#, _'CV:C# *stp.k(w)* */sṯt'pa:ku/ 'I chose'¹⁸

Table 22-3 shows the syllabic paradigms admissible in early Egyptian. Parentheses signify that the presence of the corresponding syllabic structure is not universally accepted. This is the case of the doubly-closed stressed syllable, which characterizes a certain number of plural forms of bisyllabic nouns, and of the open unstressed syllable in final position, typical for the endings of specific verbal forms (pattern -CV#) and personal pronouns (pattern -CV:C#).

Table 22-3. Early Egyptian Syllabic Structures

	Pretonic	Tonic	Posttonic
Open	#CV'_	#CV:_'	('_CV#)
Closed	#CVC'_	'CVC	'_CVC#
Doubly-closed		('CVCC#)	
Long		'CV:C#	

In historical Egyptian, the stress falls either on the ultimate (oxytone) or on

17. Schenkel 1983a: 171–230; 1990: 63–78.

18. Kammerzell 1991a: 189–92; 1991b: 198 ff. In the more traditional interpretation, the fall of final vowels is seen in connection with the transition from the *Dreisilbengesetz* to the *Zweisilbengesetz* in the prehistory of Egyptian: cf. Fecht 1960 §§ 392–406; Schenkel 1990: 78–86.

the penultimate (paroxytone) syllable of a word. Oxytone patterns¹⁹ are #CV'CVC# (*wbh* */wa'baχ/ 'to become white', Coptic ⲟⲩⲃⲁⲱ), #cvc'cvc# (*jfdw* */ja'f'daw/ 'four', Coptic ⲓⲩⲟⲟⲩⲩ), #cv:c# (*dd* */ja:d/ 'to say', Coptic ⲁⲱ), #cv'cvcc# (*mdw.w* */ma'duww/ 'words', Coptic ^B-ⲙⲧⲁⲩⲩ).²⁰ Paroxytone patterns are #cvccvc# (*stp.w* */'satpaw/ 'is chosen', Coptic ⲥⲟⲩⲧⲡ), #cv:cvc# (*stp* */'sa:tap/ 'to choose', Coptic ⲥⲱⲧⲡ), #cv'cvccvc# (*hprw.w* */χu'pirwaw/ 'transformations', Akk. transcription (*a*)*h-pí/é-ia*; Osing 1976: 558ff.), #CV'CV:CVC# (*psdw* */pi'si:jw/ 'nine', Coptic ⲩⲡⲧ), #CVC'CVCCVC# (*wpw.tjw* */wap'wutjw/ 'messengers', Akk. *ú-pu-ti/ú-pu-ut*), #cvc'cv:cvc# (*wpw.tj* */wap'wu:tj/ 'messenger', Meroitic *apote*; Osing 1976: 532–33).

Since the stress can only affect the last two syllables of an Egyptian word, the governing rule of syllabic patterns is known with the German term *Zweisilbengesetz* ('law of the two syllables'). For the prehistory of the Egyptian language, some scholars posit a situation in which, as in the related Semitic languages, the stress could also affect the antepenultimate syllable (*Dreisilbengesetz*, i.e., 'law of the three syllables'; Fecht 1960 §§ 325–47). Following the loss of the short vowel in the open posttonic syllable, words displaying this syllabic pattern were subsequently integrated into the regular patterns with penultimate stress: */χupiraw/ > */χupraw/ 'transformation'.

22.4.3. Text sample (CT II 209c–210c)

Here is a short text sample of Early Egyptian. The conventional Egyptological transliteration of the original hieroglyphic text (drawn from the funerary corpus of the Coffin Texts, about 2000 B.C.E.) is followed by a translation and a tentative phonological reconstruction.

jr.t hprw.w m bjk qj ššd snđ ntr.w
rs ʒs.t bkʒ.tj hr mtw.t sn=s wsjr
tz=s hjm.t wn.t jb=s ndm.w m mtw.t sn=s wsjr

'To make transformations as falcon. The lightning flash will strike, the gods will be afraid.

Isis will wake up pregnant from the seed of her brother Osiris.

19. In the following examples, the reconstruction of the phonological structure of a specific word in early Egyptian is accompanied by the later evidence (Akkadian transcriptions from the New Kingdom, Meroitic borrowings, or the Coptic form of the word) on which this reconstruction is based.

20. Superscript letters are abbreviations of Coptic dialects represented by the forms they precede: B, Bohairic; S, Sahidic; A, Akhmimic; L, Lycopolitan (or Subakhmimic); F, Fayyumic.

She will stand up—the hastening woman—with her heart rejoicing over the seed of her brother Osiris.’

*/ji:rit χu'pirwaw ʔm 'ba:jčk qaj 'sa:fčd ,sanja na'cu:ruw
ras 'ru:sat ba'kartčj čur ,mitwat 'sa:nčs wč'su:rčt
ca'čas 'hi:mat 'wanjat 'ji:bčs 'načmaw ʔm ,mitwat 'sa:nčs wč'su:rčt/

22.5. The phonological system of Later Egyptian (about 1000 B.C.E.)

By the end of the New Kingdom (1550–1000 B.C.E.), the phonological system described in the preceding section had undergone a certain number of developments which profoundly modified all its components. The phonology of later Egyptian is better known to us than the hypothetical reconstruction of older Egyptian thanks primarily to the cuneiform transcriptions of Egyptian words and phrases.

The major changes can be indicated as follows.

22.5.1. Consonants

(a) From the velar to the dental series, oppositions between voiced and voiceless phonemes become progressively neutralized: *t3.wj* */tarwčj/ > Akk. transcription *-ta-a-wa* ‘the Two Lands’ ~ *dbn* */di:ban/ > Akk. transcription *ti-ba-an* ‘*dbn*-weight’ (Osing 1976: 420, 619–20). (b) While palatal phonemes are regularly kept in a number of lexemes, they often tend to acquire a dental realization: *psčw* */pi:si:čaw/ > Akk. transcription *pi/e-ši-it* ‘nine’ (Schenkel 1990: 89). (c) The dental phonemes /t/ and /t/ and the glides /j/ and /w/ undergo a process of lenition to /ʔ/ at the end of a syllable, and eventually to /ø/ at the end of a word: *pč.t* */pi:čat/ > Akk. *-pi-ta* ‘bow’; *hnw* */hi:naw/ > Akk. transcription *hi-na* ‘jar’; */marjčw/ > Akk. transcription *ma-a'-ia-*, *ma-a-i-* ‘beloved’ (Osing 1976: 463, 809–10). (d) The uvular trill /R/ completes its evolution to glottal stop /ʔ/ and eventually /ø/, merging with /ʔ/ < /j/ (cf. § 22.4): while in the execration texts of the Middle Kingdom the writing <jjjčmt> renders the Sem. toponym **yarmuta* (Hoch 1991: 590), in the group writing of the New Kingdom <3> has come to indicate the *a*-vowel (ibid. 599).

22.5.2. Vowels

A series of major developments alters the vocalic system of Egyptian during the late New Kingdom, after the reign of Ramses II, i.e., from around 1200 B.C.E. onward: (a) Parallel to the so-called ‘Canaanite vowel shift’ in contemporary Northwest Semitic languages, long stressed */a:/ becomes */o:/: *hrw*

‘(the god) Horus’ */ħa:ruw/ > */ħo:rə/ (cf. the Akk. transcription of the Neo-Assyrian period *-huru-*; Fecht 1960 § 172). (b) This sound change provokes other adjustments within the system, most importantly the change of long stressed */u:/ into */e:/: *šnj* ‘tree’ */ʃu:nǝj/ > */ʃe:nə/ (cf. the Akk. transcription of the Neo-Assyrian period *-sini*; Fecht 1960 § 172; Osing 1976: 148). (c) Already in the early New Kingdom, short stressed */i/ had become */e/: cf. the Eg. anthroponym *mnj* ‘Menes’ */maʹnij/ > */maʹneʔ/ (cf. the Akk. transcription *ma-ne-e*); at a later date, probably around 1000–800 B.C.E., short stressed */e/ < */i/ and */u/ merged into the realization */e/: cf. the Eg. toponym *ḏʾn.t* ‘Tanis’ */ʃuʃnǝt/, borrowed in Hebrew at a time when the original vocalization was clearly productive (**šuʾn* > *šōʾan*), but transcribed as *še-eʾ-nu/ša-aʾ-nu* in the Neo-Assyrian period (Schenkel 1990: 87–88; Osing 1976: 377). (d) Unstressed vowels, especially in posttonic position, merged into the mid central */ə/ (the so-called shwa): *rʿw* ‘(the god) Re’ */ri:ʃuw/ > */re:ʃə/ (Akk. transcriptions *-ri-ia*, *-re-e*) ~ *nfr* ‘good’ */na:ʃir/ > */na:ʃə/ (Akk. transcription *-na-a-pa*) ~ *mṣʿ.t* */murʃat/ > */muʃə/ (Akk. transcription *-mu-a*; Osing 1976: 20, 605–6, 149). (e) A merely phonetic evolution which probably did not affect the phonological level is */i:/ > */e:/ in proximity of /ʃ/ and /j/: *wʿw* ‘soldier’ */wi:ʃiw/ (cf. the Akk. transcription from el-Amarna *ú-i-ú*) > */ʃwe:ʃə/ (cf. later transcriptions *ú-e-eh*, *ú-e-e*, *ú-e-ú*); *mḥj.t* ‘Northwind’ */maʹhi:jǝt/ > */məʹhe:ʔ/ (cf. the Akk. transcription *-ma-ḥe-e*; Osing 1976: 20–21).

Thus, we can posit the vocalic system in Table 22-4 for later Egyptian around 1000 B.C.E. While at the phonetic level the vocalic sounds have indeed evolved from the earlier system presented in § 22.4, the number of vocalic phonemes (six) remains unchanged.

Table 22-4. Later Egyptian Vowels

	Unstressed	Stressed	
		Short	Long
Front			/i:/
		/e/	
Central	/ə/		/e:/
		/a/	
Back			/o:/

22.5.3. Syllabic structures

Because of the loss of the final dentals and of the semivocalic glides caused by the strong tonic stress, the system of syllabic structures undergoes a partial reorganization, with the emergence of previously unknown or poorly documented syllabic patterns: (a) The development of '-CV:C# in plurisyllabic words (in early Egyptian, the pattern #CV:C# had a reduced functional yield, being limited to the infinitive of biradical verbs): *mḥj.t* 'the goddess Mehit' */maḥu:jǝt/ > */mǝḥu:ʔ/, cf. the Akk. transcription *-ma-ḥu-ú*, Greek *-μχης* (with */u:/ > η); *ḥmnw* 'eight' */χa'ma:nǝw/ > */χa'ma:n/, cf. el-Amarna *ḥa-ma-an* (Osing 1976: 730, 476). (b) The same development affects the pattern '-CVCC#, previously limited to some plurals of the type **maduww: z3jw.tj* 'the city of Asyut' */θǝR'jawtǝj/ > */sǝ'jawt/, cf. Neo-Assyrian cuneiform *ši-ia-a-u-ut* (Schenkel 1990: 87). (c) The fall of final consonants increases the presence of unstressed open syllables of the pattern '_CV#, which in the earlier phase of the language were limited to the endings of specific verbal forms (pattern -Cǝ#) and personal pronouns (pattern -CV:#): *ḥrj-pǝ.t* 'overseer of the troop' */ḥa:rij'pi:jat/ > */ḥəri'pi:dǝ/, cf. cuneiform *a/i/uḥ-ri-pí-ta* (ibid. 463).

Table 22-5. Later Egyptian Syllabic Structures

	Pretonic	Tonic	Posttonic
Open	(#)CV'_	'CV:_	'_CV#
Closed	(#)CVC'_	'CVC(#)	'_CVC#
Doubly-closed		'CVCC#	
Long		'CV:C#	

22.5.4. Text Sample (Wenamun 1,47–1,48)

Here is a short text sample of later Egyptian. The conventional Egyptological transliteration of the original hieratic text from the literary tale of Wenamun (about 1000 B.C.E.) is followed by the translation and a tentative phonological reconstruction.

jw dw3w ḥpr.w jw=f h3b jw=f jǝ3y=j r-ḥrj
jw p3 nṯr ḥtp.w m p3 jm w ntj-sw jm=f sp.t p3 ym

'Then morning came, and he sent and brought me up,
while the god was resting in the tent in which he was, on the shore of the sea.'

*/ʔew'dawʔə ʁaʔp ʔewʔho:ʔəb ʔewʔa'cajjvʔ ʔərha'rej
 ʔewpəʔ'no:tə hatp ʔəmpəʔʔa'mew ʔəntejsuʔa'mef ,so:pəpəʔ'jam/

22.6. The phonological system of Coptic (about 400 C.E.)

Unlike earlier stages of the language, Coptic, which is written in an alphabetic system derived from Greek, is documented in a number of closely related “dialects.”²¹ These dialects, however, do not necessarily reproduce local varieties of the language: they represent, to a large extent, discrete sets of mainly graphic conventions for rendering Egyptian in an inadequate foreign script (cf. Loprieno 1981). The two major poles of the continuum of Coptic dialects are *Sahidic*, normally considered to reflect the Theban, upper Egyptian variety of the language, documented from the 4th century C.E. and representing the language of classical Coptic literature, and *Bohairic*, the dialect of the Nile delta, documented from the 5th century C.E. and progressively established as the dialect of the liturgy of the Coptic church. Since Sahidic represents classical Coptic, it has been chosen here for the basic presentation of Coptic phonology.²² However, I shall refer to other dialects, especially Bohairic, whenever such references become necessary for the purpose of an historical or a typological analysis.

During the first millennium B.C.E. and the first centuries C.E., Egyptian continued to undergo a number of phonological changes. In the consonantal system, the tendencies described in § 22.5(a) developed further, leading to a complete neutralization of voiced plosives in the dental, palatal, and velar series: the phonemes /d/, /g/, and /z/ are present only in Greek borrowings, the rare exceptions to this rule being the result of sonorization in proximity of /n/ (e.g., $\Delta\text{N}\Gamma$ vs. ΔNOK < *jnk* ‘T’, $\Delta\text{NZH}\epsilon$ vs. $\Delta\text{NCH}\epsilon$ < ‘*t n.t sb3.w* ‘school’).

In the labial series, the situation is more complex: the voiced phoneme /b/, which by this time was probably articulated as a fricative [β],²³ is kept in all initial and medial positions ($\text{BF}\text{B}\text{OK}$ ‘servant’, ZIBOI ‘ibis’, $\text{TB}\Delta$ ‘ten thousand’), and in final position whenever it did not immediately follow the tonic vowel of a closed syllable in the earlier stages of the language, although this

21. For recent accounts and literature on Coptic dialectology cf. the corresponding entries in Atriya 1991 on Akhmimic (Nagel 1991a), Bohairic (Shisha-Halevy 1991a), Fayyumic (Kasser 1991b), Lycopolitan (Nagel 1991b), and Sahidic (Shisha-Halevy 1991b).

22. Two insightful presentations of the problems faced by linguists in the reconstruction of Coptic phonology are Satzinger 1979 and Hintze 1980.

23. Cf. its frequent alternation with <q> /f/ and <oy> /w/: SFNOYB ~ NOYq < *nbw* /na:baw/ ‘gold’, SBOINE ~ BOYWINI < *bjn.t* /bajnvt/ ‘harp’.

may indeed be synchronically the case in Coptic: **NOYB** < */na:baw/ ‘gold’. If final /b/ did follow the tonic vowel of an etymological closed syllable, whether in monosyllabic or plurisyllabic words, it became in Coptic voiceless /p/: **OYOT** < */waʕab/ ‘to be pure’, **TAT** < */dib/ ‘horn’.

Guttural fricatives of earlier Egyptian merge in Sahidic either into **ϣ** /ʃ/ (especially /χ/: e.g., **h3** ‘thousand’ */χar/ > */χaʔ/ > **ϣO**), or into **ϣ** /h/ (especially /h/ and /ç/, sometimes also /χ/: e.g., **h3.t** ‘beginning’ */ħu:rit/ > **ϣH**, **h(w).t** ‘body’ */ʕu:wat/ > **ϣH**, **hrw** ‘voice’ */χiʕraw/ > **ϣOOY**). Other dialects appear more conservative: Bohairic and Akhmimic keep the uvular fricative /χ/ (written **ϣ** in Bohairic, e.g. **ϣPWOY**, and **ϣ** in Akhmimic, e.g., **ϣPΔY** ‘voice’). Finally, the glottal stop /ʔ/, which represents the development of */ʔ/ and */ʕ/, on the one hand, and of the fall of final /t/ and /r/ on the other, is not expressed by an independent grapheme, but is rendered in Sahidic by <ϣ> at the beginning and the end of a word (e.g., **ΔNOK** /ʔaʕnok/ ‘I’ < */jaʕnak/, **TO** /toʔ/ ‘land’ < */taʔ/) and, except in Bohairic, by the reduplication of the vocalic grapheme when immediately following the stressed vowel of a word (e.g., **ϣOOOT** /χoʔp/, **ϣOOOT**, **ϣOOT** /ʃoʔp/ ‘to be’ < **hpr.w** */χapraw/ ‘has become’).²⁴

Bohairic spelling conveys a traditional feature of Egyptian phonetics, namely the aspirated realization of stops, which are expressed by the corresponding *aspiratae* of the Greek alphabet: voiceless stops become aspirated when immediately preceding a tonic vowel, semivowels, and sonorant consonants (including **B**):

$$\begin{array}{l} \text{π /p/, τ /t/, χ /c/, κ /k/} \rightarrow \text{ϕ [pʰ], θ [tʰ], ϑ [cʰ], χ [kʰ]} \\ \text{/ _'V, /b/, /m/, /n/, /l/, /r/, /w/, /j/} \end{array}$$

Examples: **SπPH** vs. **BϕPH** ‘the sun’, **STAI** vs. **BθAI** ‘this (fem.)’, **SXOEIC** vs. **BδOIC** ‘lord’, **SKOYΔAB** vs. **BXOYΔB** ‘you are holy’. This phonetic rule proves that **ϑ** [cʰ] represents in Bohairic the aspirated variety of the palatal plosive **χ** /c/; the value of the sign **ϑ** in this dialect, therefore, differs from its function in all other Coptic traditions, where it indicates the postpalatal plosive /kʰ/, a phoneme absent from the phonological inventory of Bohairic, where it has merged with **χ** /c/.

The rule of aspiration in Bohairic, however, exhibits an extremely interesting property: when /t/, /c/, and /k/ represent the outcome of older *d* /d/, *ḏ* /j/, *g* /g/, and *q* /q/ respectively, no aspiration immediately preceding the tonic

24. However, final /ʔ/ is expressed by <ϣ> in Sahidic and <ϣ> in Bohairic in doubly closed syllables, cf. below.

vowel takes place: ^{SB}ⲧⲁⲡ ‘horn’ < Eg. *db* */dib/, ^Bⲧⲱⲣⲓ parallel to ^Sⲧⲱⲣⲉ ‘hand’ < Eg. *ḏr.t* */ʔa:rat/, ^Bⲭⲓⲙⲓ ~ ^Sⲟⲓⲛⲉ ‘to find’ < Eg. *gmj.t* */gi:mit/, ^{SB}ⲕⲁⲥ ‘bone’ < Eg. *qs* */qes/; in pre-sonorant environments, on the other hand, the rule of aspiration is upheld: ^Bⲛⲉⲣⲉⲓ < *dj-jtj=f*-, ^Bⲛⲉⲃⲁ ‘ten thousand’ < *ḏb*’ /ʔa’baʃ/, ^Bⲟⲣⲏⲭⲓ ‘dowry’ < *grg.t* /ga’ru:gʊt/, ^Bⲭⲃⲟⲃ ‘to become cool’ < *qbb* /qa’bab/ (Worrell 1934: 17–23).

This phenomenon can be conveniently interpreted by assuming that in spite of the forward movement of their point of articulation which took place in later Egyptian (cf. § 22.5) from the palatal to the dental (*d* > /d/), from the velar to the palatal (*g* > /j/), and from the uvular to the velar region (*q* > /k/), these three phonemes of older Egyptian did in fact preserve their prevocalic ejective articulation down to Coptic (/j/ = [cʰ] > /d/ = [tʰ], /g/ = [kʰ] > /j/ = [cʰ], /q/ = [qʰ] > /k/ = [kʰ]); hence the use of the Greek *tenuēs* rather than of the Greek *mediae* to indicate them in the writing: ⲧ for /d/ = [tʰ], ⲭ for /j/ = [cʰ], ⲕ for /k/ = [kʰ].²⁵ On the contrary, etymological *t* /t/, *ṭ* /c/ and *k* /k/, which were not ejective but aspirated stops, maintained the aspiration under the phonetic contexts described above. Again, we can consider this aspiration graphically rendered only in Bohairic, but phonetically present in Coptic as a whole:²⁶ ^Sⲧⲁⲓ vs. ^Bⲛⲉⲃⲁ ‘spittle’ /taʃ/ = [tʰaʃ] < Eg. *tf* */tif/ = [tʰiʃ], ^Sⲧⲱⲣⲉ vs. ^Bⲛⲉⲣⲓ ‘willow’ /ʔo:rə/ = [tʰo:rə] < Eg. *trj* */ca:rʃj/ = [ʰcʰa:rʃj], ^Sⲭⲓ vs. ^Bⲟⲓ ‘to take’ /ci:/ = [cʰi:] < Eg. *ṭʒj.t* /ʰci:rʃj/ > [ʰcʰi:ʔəʃ], ^Sⲕⲏⲙⲉ vs. ^Bⲭⲏⲙⲓ ‘Egypt’ /ʰke:mə/ = [ʰkʰe:mə] < Eg. *km.t* */ku:mat/ = [ʰkʰu:mat]. This would point to a phonological, rather than phonetic status of the underlying opposition “voiceless : ejective,”²⁷ an opposition graphically conveyed only by Bohairic and displayed by the presence of minimal pairs such as ^Bⲧⲱⲣⲓ ‘hand’ < *ḏr.t* : ^Bⲛⲉⲣⲓ ‘willow’ < *trj* or ^Bⲭⲏ ‘dish’ < *ḏʒ.t* : ^Bⲟⲓ ‘quince’, of uncertain etymology.

Therefore, as in the case of its Egyptian antecedents, the phonology of Coptic dialects may actually exhibit a higher degree of complexity than is

25. Indirect evidence of the ejective character of voiceless stops in Bohairic is provided by the 13th century Arabic version of the ‘Apophtegmata Patrum’ in Coptic script: cf. Satzinger 1971: 40–65; 1991. As a rule, <ⲧ> and <ⲕ> are used in this text to render Ar. /t/ and /q/, and <ⲛ> and <ⲭ> for Ar. /t/ and /k/ respectively (in final position Ar. /t/ is sometimes rendered by <ⲧ>, and Ar. /k/ always by <ⲕ>).

26. The reason for rendering aspirated stops in dialects other than Bohairic with the corresponding Greek *tenuis* would be that Greek *aspiratae* generally represent in Coptic the combination of the corresponding voiceless phoneme followed by the glottal fricative: ⲕ = /ph/ (rather than /pʰ/), ⲛ = /th/ (rather than /tʰ/), ⲭ = /kh/ (rather than /kʰ/).

27. As generally assumed by scholars (see Kasser 1991c), except for Bohairic ⲕ, which some linguists consider phonemically distinct from ⲭ: cf. Shisha-Halevy 1991a: 54.

betrayed by a superficial graphemic analysis:²⁸ in our concrete example, we probably have to posit for the entire Coptic domain (although graphemically mirrored only in Bohairic) the presence of *three* stops in the dental, prepalatal, and velar region: (a) a voiceless series, characterized by an optional aspiration; (b) a voiced series, limited to Greek borrowings (with a few exceptions due to sonorization in proximity to /n/, cf. above); (c) an ejective series, the heir of the old voiced stops, which never exhibit aspiration and therefore resist a total fusion with the corresponding voiceless phonemes. Graphemically, the voiceless series is conveyed by the Greek *tenues* and Coptic **ⲁ** (or by the *aspiratae* or **Ⲅ** in Bohairic in stressed pre-vocalic or pre-sonorant phonetic context),²⁹ the voiced series—limited to the Greek component of the lexicon—by the Greek *mediae*, and the ejective series—limited to the Egyptian vocabulary—again by the Greek *tenues*, but this time without the Bohairic change to the corresponding *aspirata* in stressed pre-vocalic or pre-sonorant environment.

The treatment of the glottal stop /ʔ/ also deserves particular attention. As pointed out in § 22.5, in later Egyptian /t/, /tʰ/, /j/, and /w/ are dropped in final unstressed position, but become /ʔ/ when closing a stressed syllable, often representing the only remnant of an unstressed final syllable of older Egyptian that has been dropped in the later phase of the language. However, especially in final position after stressed vowels, glottal stops deriving from the development of final /t/, /tʰ/, /j/, and /w/ are not treated exactly like etymological /ʔ/; in these cases one also finds slight differences in the treatment of /ʕʔ/ < */uʔ/ as opposed to /ʕʔ/ < */iʔ/.³⁰

In the various Coptic dialects, different graphic solutions for /ʔ/ are adopted. All varieties display /ʔ/ = <ø> in initial position (cf. ^{SB}ⲁⲛⲟⲕ /ʔa'nok/, ^{ALF}ⲁⲛⲁⲕ /ʔa'nak/ < */ja'nak/ 'I'). To express a glottal stop following the tonic vowel in words of more than one syllable, all dialects except Bohairic exhibit the reduplication of the vowel's grapheme, whether the glottal stop belongs to the same syllable (the vowel being in this case short: /'CVʔ/ = <CVV>, e.g. ^Sⲧⲟⲟⲧⲩ̅, ^Bⲧⲟⲧⲩ̅ /'toʔtəf/, ^Fⲧⲁⲗⲧⲩ̅ /'taʔtəf/ < */jartʰf/ 'his hand', ^Sⲙⲟⲟⲩⲉ, ^Bⲙⲟⲩⲩ̅ /'moʔʔə/ < */maʃʃʲj/ 'to walk') or to the following

28. For the traditional assumption that Coptic, unlike any other language, displays an exact correspondence between graphemic appearance and phonological structure, cf. Kasser 1991d: 207 ff.

29. This is a general context for the development of aspiration, technically called “delayed voicing onset,” also shown, e.g., in Modern English and German; cf. Hock 1991: 121.

30. Cf. background information, discussion and examples in Osing 1976: 15–17, 403–48.

syllable (the tonic vowel being here long: /'CV:/ = <CṼV>, cf. **ΟΥΗΗΒ** /'we:ʔəb/ < */wi:ʕv̄b/ 'priest'). In this last case, i.e., if /ʔ/ is the first phoneme of a final syllable of the type -ʔVC# following a stressed syllable of the type #CV:,³¹ it is conveyed in most dialects by the reduplication of the tonic vowel, and in Bohairic by <ø>: ^S**ΛΩΩΜΕ**, ^B**ΛΩΜ** /'co:ʔəm/ < */ja:miʕ/ 'book'. That the interpretation of the phonological structure presented here is plausible is shown by two facts: (a) The interesting graphemic opposition found in Bohairic between the writing <-Cɿ> to express final /-Cə/, as in ^B**ΡΩΜΙ** /'ro:mə/ 'man' or ^B**ΜΩΨΙ** /'moʔʕə/ 'to walk', and the writing <-øC> to express final /-ʔəC/, as in ^B**ΜΗΨ** 'crowd', whereas in Sahidic both environments are graphically rendered by <-Ce>: ^S**ΡΩΜΕ**, ^S**ΜΟΟΨΕ**, ^S**ΜΗΗΨΕ**. (b) The two graphic renditions exhibited by the unstressed syllabic structure /-ʔəC#/ in Sahidic, namely <-VVCe> as in **ΛΩΩΜΕ** /'jo:ʔəm/, but also <-VVC> as in **ΒΩΩΝ** /'bo:ʔən/. There can be no doubt that these two patterns are phonologically identical: cf. on the one hand the Sahidic variant with final -ε (^S**ΒΩΩΝΕ**), on the other hand the identical treatment of these two structures in the other dialects: cf. ^A**ΛΟΥΟΥΜΕ**, ^B**ΟΥΟΥΝΕ**, ^B**ΛΩΜ**, **ΒΩΝ**.

If /ʔ/ represents the last phoneme of a doubly-closed final syllable, it is expressed in the writing by <ε> in the dialects of Upper Egypt and by <ɿ> in those of Lower Egypt: ^S**ΕΙΟΤΕ**, ^B**ΙΟΤ** /jotʔ/, ^{AL}**ΕΙΑΤΕ**, ^{ALF}**ΕΙΑΤ** /jatʔ/ < */jatjaw/ 'fathers', ^S**ΖΙΟΜΕ**, ^B**ΖΙΟΜΙ** /hjomʔ/, ^{AL}**ΖΙΑΜΕ**, ^F**ΖΙΑΜΙ** /hjamʔ/ < */hi'jamwv̄t/ 'women'.³²

A last problem is represented by the fate of the phoneme /ʕ/. Its existence, although not excluded, is in fact very doubtful. The graphic distribution of etymological /ʕ/ is identical with that of etymological /ʔ/, including /ʔ/ < /j/, /w/, /r/, and /t/, and scholars generally maintain³³ that it had already merged with the glottal stop /ʔ/ in later pre-Coptic Egyptian, but left traces in Coptic vocalism, especially in the anteriorization of its vocalic surrounding: unstressed **Δ** instead of **ε** or <ø> (as in **ΔΨΔΙ** < ʕʕ * /ʕi'ʕir/ > * /ʕə'ʕiʔ/ 'to become numerous' vs. **ϸΖΔΙ** < ʕʕʕ * /θi'ʕir/ > * /sə'ʕiʔ/ 'to write'), stressed **Δ** instead of

31. In most of these words, the /ʔ/ derives from an etymological /ʕ/ via metathesis: ^S**ΤΩΩΒΕ**, ^B**ΤΩΒ/ΤΩΠ** /'to:ʔəb/ < */ja:baʕ/ 'to seal'.

32. Many scholars would interpret the syllabic structure of these words somewhat differently, namely as ^S**ΕΙΟΤΕ**, ^B**ΙΟΤ** = /'jotə/. From the point of view of the economy of a linguistic system, however, this phonological analysis presents the drawback of positing the existence of a stressed open syllable /'CV-/ in a polysyllabic word, which is not documented throughout the history of the Egyptian language and is wholly unnecessary at the purely synchronic level as well: cf. § 22.6.2 and n. 37.

33. Cf. the bibliographic information in Kasser 1991a.

o (as in **ἡ** < *db* */ja'baʃ/ > */tə'baʃ/ 'ten thousand' vs. **κροϥ** < *qrf* */qa'raf/ > */qə'raf/ 'ambush').³⁴

22.6.1. Vowels

Table 22-7 presents the vocalic system of Sahidic Coptic around 400 C.E.

Table 22-7. Sahidic Coptic Vowels

		Stressed	
		Short	Long
		<(ϵ)ɪ> /i:/	
Front	<ɪ> /i/	<ϵ>, <ø> /e/ ^a	
		<ɥ> /e:/	
Central	<ϵ>, <ø> /ə/	<ɔ> /a/	
		<ω> /o:/	
Back	<ɔ> /a/	<o> /o/	
		<oʏ> /u:/	

^aAs we saw above, /e/ = <ø> in Sahidic, Akhmimic, and Lycopolitan, <ϵ> in Bohairic, and <ɥ> or <ʏ> in Fayyumic before sonorant phonemes (including **ʁ**).

When compared with the preceding phases in the history of Egyptian, the vocalic system of Coptic exhibits the continuation of the later Egyptian *Lautverschiebung*. Later Eg. stressed */a/ becomes /o/ in the two major dialects (Eg. *sn* */san/ 'brother' > ^{SB}**CON**, ^{ALF}**CAN**), following the pattern of the evolution */a:/ > /o:/ (Eg. *rm̄* */ra:mac/ 'man' > */ro:mə/ > **ρOME**) which had already taken place around 1000 B.C.E. (cf. § 22.5). Moreover, Later Eg. stressed */e/, whether deriving from original */i/ (as in *m* */rin/ > */ren/ 'name') or from original */u/ (as in *hrw* */χurraw/ 'Hurrian' > */χel/ 'servant'), becomes /a/ in Sahidic and Bohairic, whereas it is kept as /e/ in the minor dialects: ^{SB}**PAN**, ^{AL}**PEN**, ^F**LEN**; ^{SL}**2AL**, ^A**2EL**, ^F**2EL**.

These two developments in the quality of the short stressed vowels display a certain number of exceptions, of phonetic (if not purely graphemic) rather than phonological character and generally motivated by specific consonantal surroundings. Thus, */a/ is kept as /a/ in the two major dialects and is

34. For other possible signals of a preservation of the phoneme /i/ in final position cf. the discussion on the glottal stop /ʔ/ in § 22.6.2.

rendered as <ε> in Fayyumic before etymological guttural fricatives (^{SAL}ⲧⲃⲁ, ^Bⲉⲃⲁ, ^Fⲧⲃⲉ < ḏb·*/ja'baʃ/ 'ten thousand'); conversely, */a/ becomes /o/ also in Akhmimic and Lycopolitan before etymological /ʔ/ and /ʕ/ (^Sⲉⲓⲟⲟⲣ(ε), ^Bⲓⲟⲣ, ^Aⲓⲟⲟⲣⲉ, ⲓⲱⲱⲣⲉ, ^Fⲓⲁⲁⲗ, ⲓⲁⲁⲣ < jtrw */jatraw/ > */jaʔr(ə)/ 'river'). Also, the diphthongs */aj/ and */aw/, which regularly yield /oj/, /ow/ in Sahidic and /aj/, /aw/ in the minor dialects, appear written in Bohairic as <ⲱⲓ> (except in final position) and <ⲱⲟϣ> (in all positions) respectively: ^Sⲉⲣⲟⲓ, ⲉⲣⲟⲟϣ, ^{AL}ⲁⲣⲁⲓ, ⲁⲣⲁϣ, ^Fⲉⲗⲁⲓ, ⲉⲗⲁϣ, ^Bⲉⲣⲟⲓ, ⲉⲣⲱⲟϣ 'to me, to them'.

As for */e/, which, as we saw, regularly turns into ^{SB}ⲁ and ^{ALF}ⲉ, its main phonetic (or graphemic) exception is represented by its being written as <ⲟ>³⁵ in Sahidic, Akhmimic, and Lycopolitan, as <ε> in Bohairic, and as <ⲙ> or <ϣ> in Fayyumic, before sonorant phonemes (including ⲃ): *šmsj* */fimsij/ > ^{SAL}ⲱⲙⲱⲉ, ^Bⲱⲉⲙⲱⲓ, ^Fⲱⲙⲱⲓ /femʃə/ 'to worship'. In all dialects except Bohairic, if the following sonorant is not followed by another consonant, it is subject to reduplication: *qnj.t* */qinjɪt/ > ^Sⲕⲏⲏⲉ, ^Aⲕⲏⲏⲓⲉ, ^Bⲕⲉⲏⲓ, ^Fⲕⲏⲏⲏⲓ 'to become fat'. Also, in proximity of sibilants one may find the outcome */e/ > ^{SB}ⲉ or even ^{SBAF}ⲓ: e.g., *wš.t* */wisχat/ > ^Sⲟϣⲉⲱⲉ, ^Sⲟϣⲟⲱⲉ, ^Bⲟϣⲉ/ⲙⲱⲉ 'breadth', *pš.s.t* */puʃsat/ > ^Sⲡⲱⲉ, ^πⲁⲱⲉ 'half'. The diphthongs display slight irregularities as well: instead of the paradigmatic form <ⲁϣ> (as in *snwj* */si'nnewwɔj/ > ^Sⲕⲏⲁϣ 'two', *hnw* */hṽnew/ > ^Sⲕⲏⲁϣ 'jar'), */ew/ occasionally yields <ⲟϣ>, and <ⲟ> in Akhmimic in final position: ^Sⲕⲏⲟϣ, ^Aⲕⲏⲟ, ^Sⲕⲏⲟϣ. The outcome of */ej/ is even more complex: it develops as expected into ^{SL}ⲁ(ε)ⲓ, but it keeps a vocalization closer to the original in ^Aⲉ(ε)ⲓ, ^Fⲙⲓ; Bohairic exhibits a difference in treatment, depending on whether the original vowel was *u (i.e., */ej/ < */uj/), in which case it goes with Sahidic ⲁⲓ, or *i (i.e., */ej/ < */ij/), in which case it goes with Fayyumic ⲙⲓ: e.g., *zjnw* */θijnṽw/ > ^Sⲕⲁⲉⲓⲏ, ^Aⲕⲉ(ε)ⲓⲏⲉ, ^{BF}ⲕⲏⲓⲏⲓ 'physician', *jḳy* */ʕujqṽj/ > ^{SL}ⲁⲉⲓⲕ, ^Bⲁⲓⲕ 'consecration'.

Moving on to the long vowels, Coptic displays no major phonological development from the later Egyptian system. At the phonetic level, the following phenomena take place: (a) All dialects exhibit an evolution */a:/ > <ⲟϣ> [u:] (instead of */a:/ > /o:/), regularly after nasal consonants, and occasionally following other consonants as well: *ntr* */na:car/ > ^Sⲏⲟϣⲧⲉ /nu:te/ 'god'. Akhmimic also displays <ⲟϣ> in final position or if the vowel is followed by the glottal stop, i.e., by a reduplication of the vocalic grapheme: ^Sⲁⲱⲱⲙⲉ,

35. The presence of a short, non-phonemic vowel [ə] is indicated in most dialects by a supralinear stroke (called in German *Vokalstrich*) over the following consonant.

^Aⲭⲟⲩⲟⲩⲙⲉ. That /u:/, nonetheless, has acquired phonemic character in Coptic is shown by the presence of minimal pairs such as ⲫⲱⲛ /ho:n/ < *hnn* */'ça:nan/ 'to approach' vs. ⲫⲟⲩⲛ /hu:n/ < *hnnw* */'ça:naw/ 'inside'. (b) Wide-spread is also the outcome <(ε)ɪ> [i:] instead of /e:/ from etymological */u:/ > */e:/ (cf. § 22.5) in proximity of /r/ and after etymological pharyngeals: ^{SL}ⲫⲓⲣ, ^Bⲫⲓⲣ, ^Aⲫⲓⲣ, ^Fⲫⲓⲣ < */'χu:r/ 'street' (Sem. loanword). As in the case of */a:/ > <ⲟⲩ> [u:], here too Akhmimic displays <εɪ> in final position or if the vowel is followed by /ʔ/: ^Sⲧⲏⲏⲃⲉ, ^Aⲧⲉⲓⲃⲉ 'finger'. This same */u:/ > */e:/ occasionally appears as <ε> before pharyngeal phonemes: ^Sⲭⲙⲡⲉⲗ < */tap'pu:h/ 'apple' (Sem. loanword). (c) We had already observed in later Egyptian the phonetic outcome */i:/ > */e:/ in proximity of /ʕ/ or /j/ (cf. § 22.5).

Most Coptic dialects have three unstressed vocalic phonemes (cf. Osing 1976: 27–30, 475–500), depending on the phonetic context of the original structure of the word: as a general rule, pretonic and posttonic vowels have developed into /ə/, ³⁶ graphically rendered as <ε> or <ø> (<ɪ> in the northern dialects Bohairic and Fayyumic in final position); pretonic unstressed /i/, which often alternates with <ø>, derives from a pretonic unstressed syllable of the type */CVj-/ (^Sⲫⲓⲃⲱⲓ 'ibis' < *h(j)bj.w* */hij'ba:jvw/, originally the plural of *h(j)bw* */hijbaw/ > */hi:b/, cf. ^Bⲫⲓⲡ), whereas pretonic unstressed /a/ owes its origin to an earlier Egyptian unstressed */a/ (original or resulting from assimilation of */e/ < */i/ or */u/ in proximity of an etymological pharyngeal or velar phoneme: ⲭⲱⲁⲓ 'to become numerous' < šʕ */ʕi'ʕiʕiʕ/, or to an unstressed sonorant phonetic surrounding (ⲭⲙⲡⲏⲫⲉ 'asphalt' < */mʕ'rihjat/).

22.6.2. Syllabic structures

Coptic syllabic patterns are very similar to those of later Egyptian, the only major difference being represented by the emergence of new patterns from the reduction to shwa and eventually to zero of the short vowel of pretonic open syllables: */#CV-CV(C) > #C-CV(C). As in the earlier stages of the language, long and doubly-closed syllables are documented only in stressed final position.

These rules of syllabic distribution and the following comments apply to the vocabulary of Egyptian stock, not to the Greek words which entered the language especially in the religious sphere of vocabulary.

36. If the stressed syllable of earlier Egyptian was of the type CV:- and the first consonant of the posttonic syllable was /w/, /j/, or /ʔ/, Egyptian posttonic vowels in syllables of the type -CVw, -CVj, and -CVʔ have left different traces in the final long vowels or diphthongs of Coptic (Schenkel 1990: 91f.).

Table 22-8. Sahidic Coptic Syllabic Structures

	Pretonic	Tonic	Posttonic
Open	(#)CV'_	'CV:(#)	'_CV#
	#CCV'_	'CCV:(#)	
Closed	(#)CVC'_	'CVC(#)	'_CVC#
	#CCVC'_	'CCVC(#)	
Doubly-closed	(#)CVCC'_	'CVCC#	
	#CCVCC'_	'CCVCC#	
Long		'CV:C#	
		'CCV:C#	

A pattern of tonic open syllable with short vowel ('CV) is apparently documented in words such as **πε** 'heaven' < *p.t* */pit/, **το** 'land' < *tʒ* */tar/, **ϣαλε** 'to tell' < *sdd.t* */sijdit/, or **ειοπε** 'occupation' < *wpw.t* */japwat/. However, I adopt here a more economic approach to Coptic phonology, which displays the supplementary advantage of establishing a continuity between earlier Egyptian and Coptic. It consists in analyzing this syllabic pattern as closed ('CVC) or doubly closed ('CVCC), by positing the existence of a final glottal stop /ʔ/: thus **πε** = /peʔ/, **το** = /toʔ/, **ϣαλε** = /ʃaʔ/, and **ειοπε** = /jopʔ/, parallel to the pattern **ραν** = /ran/ or **κοτπ** = /sotp/ 'chosen'.³⁷

Two important elements in favor of this analysis are: (a) the graphic rendering of this glottal phoneme as final <-ε> (in Akhmimic and Lycopolitan) or <-ι> (in Bohairic and Fayyumic) in dialects other than Sahidic, and occasionally in Sahidic itself: cf. ^SME, MEε, ^{SAL}MHE, ^AMIε, ^BMHI, MEI, ^FMEI, MEEI, MHI 'truth' = in all cases /mVʔ/; (b) the Akhmimic (and partially Lycopolitan) treatment of etymological */a/ as <O> (or sometimes <ε>) rather than <Δ> and of etymological */a:/ as <OY> rather than <ω> in final position and before reduplication of the vowel (^{SAL}TOOTġ, ^BTOTġ, ^FTAΔTġ 'his hand'; ^{SBAL}NTO, ^FNTΔ 'you (fem.)', ^{SF}KW, ^BχW, ^LKW(ε), ^{AL}KOY 'to lay'; ^SΔΩΩME, ^AΔOYOYME 'book'), two environments which were evidently perceived to share common features. This phenomenon points to the fact that what appears here graphically to be a final vowel is in fact a phonological /ʔ/. It also needs to be stressed that this glottal stop is always justified at the

37. Needless to say, the phonetic realization of these phonological strings may very well have been [ʃaʔə], [jopə], or [sotəp], but in this instance the phonetic dimension is both impossible to reconstruct and irrelevant within the context of our discussion.

etymological level (coming from the lenition and eventually the fall of an earlier consonant, cf. § 22.5).

That this final glottal stop is not expressed in the writing should hardly be surprising, since this is the regular fate of /ʔ/ in Coptic in all initial and final positions, unless it represents the last phoneme of a doubly-closed syllable of the type we considered above (ϵιοⲡⲉ = /jopʔ/). Accordingly, a structure such as ⲧⲟⲉ ‘part’ < *dnj.t* */danjut/ (cf. Osing 1976: 440) should probably be analyzed as /toʔʔ/, the sequence of two glottal stops at the end of the doubly-closed syllable being the reason for the variety of writings of this word (ⲧⲟⲉ, ⲧⲁ(ϵ), ⲧⲟ, to mention just the Sahidic forms).

Conversely, the apparent and utterly un-Egyptian presence of patterns with long unstressed vowel (CV:'_ as in **ΟΥΤΑΞ** 'fruit' or '_CV: as in **ΤΗΡΟΥ** 'all of them') is easily removed from the phonological system of Coptic by interpreting <ΟΥ> in these cases as /w/: **ΟΥΤΑΞ** = /wtah/, pattern 'CCVC and **ΤΗΡΟΥ** = /'te:rəw/, stressed pattern 'CV: and posttonic pattern '_CVC#. In both cases, the hypothetical [u:] in *[u:'tah] or *[te:ru:] represents the realization of /w-/ and /-əw/ respectively in these specific phonetic contexts.

22.6.3. Text sample

Here is a short passage from the works of Shenute (4th century C.E.), one of the main figures of the Coptic church and one of the classical authors of Coptic literature, after Till 1970: 293–94. The Coptic text is followed by the conventional scholarly transliteration. Greek borrowings are transliterated and transcribed in italiics.

ΝΨΩΧΑΠ̅ Ν̅ΝΨΔΔΕ Μ̅ΠΕΙΔΩΜΕ Η ΠΚΕΣΕΕΠΕ ΕΝΤΑΝΧΟΟΥ ΔΥΩ
ΔΝCΔΖΟΥ Ζ̅ΝΤΜΕΖΡΟΜΠΕ C̅ΝΤΕ Μ̅Ν̅ΝCΔ ΤΡΕΝΚΩΤ Μ̅ΠΙΗ Ζ̅Μ̅ΠΚΑΙΡΟC
ΕΝΤΑ̅ΝΒΑΡΒΑΡΟC ΨΩΛ ΨΔΝΤΟΥΒΩΚ ΕΖΟΥΝ ΕΤΠΟΛΙC ΕΤΟΥΜΟΥΤΕ ΕΡΟC
ΔΕ ΚΟΕΙC Ζ̅Μ̅ΠCΗΥ ΕΝΤΑΠΕΙΝΟΒ̅ Μ̅ΜΗΗΨΕ ΔΟΕΙΛΕ ΕΡΟΝ

<nšōjp nnšaje mpeiĵōōme ē pkeseepe entanjoou auō ansahou
hntmehrompe snte mnnsa trenkōt mpiēi hmp*kaïros*
entan*barbaros* šōl šantoubōk ehoun et*polis* etoumoute eros
je koeis hmpsēu entapeinoc mmēēše coeile eron>

/ʔən'fo:ɕəp ʔənʔən'jaɕʔ ʔəmpəj'jo:ʔəm e: pɕə'seʔpə ʔəntən'joʔw ʔa'wo:
ʔən'sahw həntməh'rompə 'səntə mənʔən'saʔ trən'ko:t ʔəmpi'ʔe:j həmp 'kajros
ʔəntəʔən 'barbaros ʔo:l fantəw'bo:k ʔə'hu:n ʔet'polis ʔetəw'mu:tə ʔə'ros
ʔə-'kojs həmp'se:w ʔəntəpəj'nokʔ ʔəm'me:ʔəʔ 'k'ojlə ʔə'ron/

‘The rest of the words of this book, i.e., the remaining (of the things) which we said and wrote in the second year after we built this house, at the time when the Barbarians plundered, until they reached the city called Qus, at the time when this huge crowd stayed with us.’

22.7. Further reading

The most accessible introduction to the study of Egyptian phonology is offered by Schenkel 1990, where the reader will find a history of the scholarly endeavors involving Egyptian phonological reconstruction, a description of the different methodological approaches, and a presentation of the contemporary state of the art, covering the Afroasiatic background, pre-Coptic phonology, and the fundamental rules of phonological development from older Egyptian to Coptic.

The most complete reference book for the study of vocalism throughout the history of Egyptian is Osing 1976, which systematizes the approach inaugurated by Fecht 1960. Two works by Schenkel were conceived on the footsteps of Osing’s treatise, completing it and revising some of its assumptions: 1983a, 1983b. Of major significance for the study of Egyptian syllabic orthography and of the phonological correspondences between Egyptian and Semitic is the doctoral dissertation by J. E. Hoch (1991).

A radically different paradigm from the one followed by Fecht, Osing, and Schenkel has been pursued by scholars who assumed a much higher degree of correspondences between Egyptian and Semitic phonology and especially morphology: in this tradition cf. Vergote 1973–83 and Vycichl 1990.

A detailed bibliography on Egyptian and Coptic phonology can be derived from Osing 1976, Schenkel 1990, and Hoch 1991.

References

- Atiya, Aziz S., ed. 1991. *The Coptic Encyclopedia*, vol. 8. New York: Macmillan.
- Baldi, Philip, ed. 1990. *Linguistic Change and Reconstruction Methodology* (Trends in Linguistics, Studies and Monographs 45). Berlin: Mouton de Gruyter.
- Brockelmann, Carl. 1908. *Grundriß der vergleichenden Grammatik der semitischen Sprachen*, vol. 1. Berlin: Reuther & Reichard.

- Eyre, Christopher, and John Baines. 1989. "Interactions between Orality and Literacy in Ancient Egypt." In *Literacy and Society*, ed. Karen Schousboe and M. T. Larsen, pp. 91–119. Copenhagen: Akademisk Forlag.
- Faber, Alice. 1990. "Interpretation of Orthographic Forms." In Baldi 1990: 619–37.
- . 1992. "Second Harvest: *šibbōleθ* Revisited (Yet Again)." *Journal of Semitic Studies* 37: 1–10.
- Fecht, Gerhard. 1960. *Wortakzent und Silbenstruktur: Untersuchungen zur Geschichte der ägyptischen Sprache* (Ägyptologische Forschungen 21). Glückstadt: Augustin.
- Hintze, Fritz. 1980. "Zur koptischen Phonologie." *Enchoria* 10: 23–91.
- Hoch, James E. 1991. "Semitic Words in Egyptian Texts of the New Kingdom and Third Intermediate Period." Ph.D. dissertation, University of Toronto.
- Hock, Hans Heinrich. 1991. *Principles of Historical Linguistics*, 2nd ed. Berlin: de Gruyter.
- Kammerzell, Frank. 1991a. "Augment, Stamm und Endung: Zur morphologischen Entwicklung der Stativkonjugation." *Lingua Aegyptia* 1: 165–99.
- . 1991b. "Personalpronomina und Personalendungen im Altägyptischen." In *Ägypten im afro-orientalischen Kontext: Gedenkschrift Peter Behrens*, ed. Daniela Mendel and Ulrike Claudi, pp. 177–203 (Afrikanistische Arbeitspapiere, Sondernummer). Cologne: Institut für Afrikanistik.
- . 1992. Review of *Les langues dans le monde ancien et moderne*, troisième partie, *Les langues chamito-sémitiques*. *Lingua Aegyptia* 2: 157–75.
- Kasser, Rodolphe. 1991a. "Ayin." In Atiya 1991: 45–47.
- . 1991b. "Fayyumic." In Atiya 1991: 124–31.
- . 1991c. "Phonology." In Atiya 1991: 184–86.
- . 1991d. "Syllabication." In Atiya 1991: 207–14.
- Lambdin, Thomas O. 1952. *Egyptian Loanwords and Transcriptions in the Ancient Semitic Languages*. Baltimore: Johns Hopkins University Press.
- Loprieno, Antonio. 1981. "Methodologische Anmerkungen zur Rolle der Dialekte in der ägyptischen Sprachentwicklung." *Göttinger Miszellen* 53: 55–75.

- . 1986. *Das Verbalssystem im Ägyptischen und im Semitischen: Zur Grundlegung einer Aspekttheorie* (Göttinger Orientforschungen 4/17). Wiesbaden: Harrassowitz.
- Nagel, Peter. 1991a. "Akhmimic." In Atiya 1991: 19–27.
- . 1991b. "Lycopolitan." In Atiya 1991: 151–59.
- Osing, Jürgen. 1976. *Die Nominalbildung des Ägyptischen*, 2 vols. Mainz: von Zabern.
- Rössler, Otto. 1971. "Das Ägyptische als semitische Sprache." In *Christentum am Roten Meer*, ed. Franz Altheim and Ruth Stiehl, vol. 1, pp. 263–326. Berlin: de Gruyter.
- Satzinger, Helmut. 1971. "Zur Phonetik des Bohairischen und des Ägyptisch-Arabischen im Mittelalter." *Wiener Zeitschrift für die Kunde des Morgenlandes* 63–64: 40–65.
- . 1979. "Phonologie des koptischen Verbs (sa'idischer Dialekt)." In *Festschrift Elmar Edel*, ed. Manfred Görg, pp. 343–68 (Ägypten und Altes Testament 1). Wiesbaden: Harrassowitz.
- . 1991. "Pronunciation of Late Bohairic." In Atiya 1991: 60–65.
- Schenkel, Wolfgang. 1983a. *Aus der Arbeit an einer Konkordanz zu den alt-ägyptischen Sargtexten*, vol. 2: *Zur Pluralbildung des Ägyptischen* (Göttinger Orientforschungen 4/12). Wiesbaden: Harrassowitz.
- . 1983b. *Zur Rekonstruktion der deverbalen Nominalbildung des Ägyptischen* (Göttinger Orientforschungen 4/13). Wiesbaden: Harrassowitz.
- . 1986. "Das Wort für 'König (von Oberägypten)'." *Göttinger Miscellen* 94: 57–73.
- . 1990. *Einführung in die altägyptische Sprachwissenschaft* (Orientalistische Einführungen). Darmstadt: Wissenschaftliche Buchgesellschaft.
- Schmalstieg, William R. 1990. "A Few Issues of Contemporary Indo-European Linguistics." In Baldi 1990: 359–74.
- Shisha-Halevy, Ariel. 1991a. "Bohairic." In Atiya 1991: 53–60.
- . 1991b. "Sahidic." In Atiya 1991: 194–202.
- Till, W. C. 1970. *Koptische Grammatik (Säidischer Dialekt)*, 2nd ed. Leipzig: Verlag Enzyklopädie.
- Vergote, Jozef. 1973–83. *Grammaire copte*, 2 vols. Louvain: Peeters.
- von Soden, Wolfram. 1969. *Grundriss der akkadischen Grammatik* (Analecta Orientalia 33/47). Rome: Pontifical Biblical Institute.

- Vycichl, Werner. 1990. *La vocalisation de la langue égyptienne*, vol. 1: *La phonétique* (Bibliothèque d'Étude 16). Cairo: Institut Français d'Archéologie Orientale.
- Westendorf, Wolfhart. 1965. *Koptisches Handwörterbuch*. Heidelberg: Carl Winter Universitätsverlag.
- Worrell, William H. 1934. *Coptic Sounds* (University of Michigan Studies, Humanistic Series 26). Ann Arbor: University of Michigan Press.
- Zeidler, Jürgen. 1992. Review of Petráček, *Vergleichende Studien*. *Lingua Aegyptia* 2: 189–222.

Berber Phonology

Maarten G. Kossmann and Harry J. Stroomer

Rijksuniversiteit te Leiden, The Netherlands

23.1. Berber languages

The Berber languages, spoken in North Africa by some 15 to 20 million people, are a branch of the Afro-Asiatic phylum.

The largest population of Berberophones can be found in Morocco. Approximately 45% of the total Moroccan population (26 million) speaks a Berber language as a mother tongue. People of the High Atlas and the Anti-Atlas mountains and the Sous valley speak Tashelhit (*tašlḥiyt*), also called Sous Berber (*tasusit*). This language is spoken by some 7 million. It is, together with Kabyle in Algeria, the most important Berber language in terms of number of speakers. Tashelhit Berber has some dialect variation, but not so strong as the Berber language spoken in the Middle Atlas mountains. This language, often called Tamazight (*tamaziyt*), is spoken by some 3 million people in various dialects. Strong dialect variation can also be found in the Moroccan Rif mountains. The Riffian language (*θarifəçθ*) has some 2 million speakers.

In Algeria, 25% of the total population (about 26 million) speaks a Berber language. In a densely populated area in the north of Algeria, Kabyle Berber (*θaqbayliθ*) is spoken by approximately 7 million. In Algeria, Berber is also spoken in the Aurès mountains, the Mzab region, the Ouargla oasis, and by the sedentary population of the Sud Oranais area. Touareg, a Berber language spoken by a million, is found not only in the Algerian Sahara but also in neighboring areas of the Sahel republics of Mali and Niger. The Tamahaq (*tamaahaq*) dialect is spoken in the Ahaggar region in southern Algeria. The Tamajaq (*tamaažaq*) dialect is spoken in the Ayr region of Niger. The Tamašek dialect (*tamašəq*) is spoken in the Adrar des Ifoghas region in Mali. The Tawlemmet dialect (*tawləmmət*) is spoken by the Iwlemmeden Touaregs of the Mali–Niger borderland.

In Tunisia, Berber is spoken by the population of fewer than six villages on the Tunisian mainland. On the island of Djerba one finds some five

Berberophone villages. All together, Berber represents only 1% of the total population.

In Libya, some 25% of the total population of 4 million is Berberophone. In this country Berber is spoken in Zouara on the western Libyan coast, in the western Djebel Nefousa region, and in a number of oases.

As far east as Egypt we find a very small community of Berberophone people in the famous oasis of Siwa. In Mauritania a small group (say, around 3,000) of Zenaga Berber speakers still exists south of the capital Nouakchott.

Berbers emigrated from their North African homelands in all directions. One finds a considerable number of speakers not only in the big cities of Morocco and Algeria but also in several European countries.

23.2. Writing

The oldest epigraphic records that perhaps represent a form of Berber are the so-called Libyco-Berber inscriptions. They stem from the pre-Islamic period and are found mainly in Tunisia and Algeria, but also in Morocco. These inscriptions are undated and difficult to interpret. The script in which they were written resembles the so-called Tifinagh script, still in use among the Touaregs. There is some regional variety in the shape of the characters. The Arabic script is also used for writing Berber. There is a longstanding tradition of writing Tashelhit in Arabic characters. The Roman script is used for scientific and practical purposes.

23.3. Classification of Berber languages

Berber languages may be tentatively classified according to morphological criteria as follows:

- Group 1. Tashelhit and Middle Atlas Berber (with exception of the Beni Ouarayn Berber and Ayt Seghrouchen of the Eastern Middle Atlas)
- Group 2. Zenati languages: Beni Ouarayn, Ayt Seghrouchen, Rif, Chawia, the dialects of the Sud-Oranais (including Figuig), Mzab, Ouargli
- Group 3. Kabyle
- Group 4. Touareg, Ghadamsi
- Group 5. Zenaga of Mauritania

23.4. Phonology

Most grammars and textbooks pay but little attention to phonology. However, during the last two decades under the influence of modern trends in theoretical phonology, scholarly interest in the phonology of Berber languages has increased.

The variation in Berber phonological systems is large. Therefore, it is not possible to give an overview of Berber phonology without referring to particular languages (see §§ 23.6–8).

23.4.1. Vowels

Most Berber languages have *a*, *i*, and *u*. According to the context, the phonetic realizations of these three vowels may be quite different, *a* ranging from [a] to [æ], *i* from [e] to [i], *u* from [o] to [u]. There are no diphthongs or vowel clusters. In some languages (Touareg, Ghadamsi, Zenaga), the vowel system is more complicated. The Touareg system is given in § 23.8.

23.4.2. The shwa

A problem in Berber phonology is the status of shwa or ə. Most Berber words contain ə or one or more syllabic consonants.

For some dialects, the leading principle for the placement of ə is the structure of the word. The principle of ə-placement is simple: If there is a cluster of two consonants, ə is placed between them, unless this would lead to ə in an open syllable (i.e., it would be followed by a single consonant followed by ə or a plain vowel). The rule operates from right to left. E.g.:

ilm → *iləm* ‘skin’ (Figuig)

išrz → *išrəz* ‘he cultivated’ (Figuig)

In *išrz*, the cluster *šr* cannot be broken by ə, as this would lead to ə in an open syllable: **išərəz* is an impossible string.

For other dialects, the leading principle for the placement of ə is the intrinsic sonority of the consonant. Consonants are ranked on a sonority scale, in which some consonants are more apt to be preceded by ə than others. The placement of ə in a string of consonants follows this scale. First, ə is placed before the consonants which are highest on the sonority scale, then before the second-highest consonants, etc. Again, ə in an open syllable is impossible. For instance, in Tashelhit, *r* ranks higher on the sonority scale than *k*, *s*, or *z*:

ikrz → *ikərz* ‘he plowed’

iskr → *iskər* ‘he did’

The prohibition of ə in an open syllable renders *ikərəz or *isəkər impossible.

These two principles account for the majority of cases where ə is found. However, almost all Berber languages have words or morphological contexts where the rules are violated. This fact has led to positing two types of ə: one inserted by phonetic rule, the other as a part of the underlying structure. This underlying ə is phonemic. It remains to be seen which of the Berber languages require an underlying ə. Its existence is certain in some of them, for instance in Figuig Berber.

23.4.3. *Consonants*

Almost every Berber language has bilabial, dental, palatal, velar, uvular, pharyngeal, and laryngeal consonants, and many have interdentals as well. A large number of consonants have been borrowed from Arabic or European languages, e.g., *ʃ*, *t*, *q*, *h*, and *ʕ* from Arabic and *p* from Spanish or French. Due to the massive influx of foreign vocabulary in some Berber languages, these borrowed consonants can be frequent.

Berber consonant systems can be described by a number of correlations. The Berber languages have voiced and voiceless consonants and fricatives, plosives, and approximants. Two correlations will be treated here: the contrast between pharyngealized and non-pharyngealized consonants, and that between lax and tense ones.

23.4.3.1. *Pharyngealization*

Berber languages contrast non-pharyngealized and pharyngealized consonants. They may be voiced or voiceless (e.g., *ɖ* and *ʈ*) and fricative or plosive (e.g., *ɗ* and *ʒ*). With the exception of some dialectal phonemes of marginal functionality, all pharyngealized phonemes are dental or interdental. Parallel to the situation in many Arabic dialects, pharyngealization is a spreading feature. Under the influence of a pharyngealized phoneme, other sounds may become pharyngealized. This process is automatic, and the pharyngealization of the other sounds is not phonemic. The domain of spreading is not the same in all varieties of Berber. Some languages seem to have the syllable or the word as a domain, while others have more complicated rules.

Proto-Berber had only two pharyngealized phonemes, *ɖ* and *ʒ*. The other pharyngealized phonemes were borrowed from Arabic. Assimilations may sometimes lead to new pharyngealized phonemes. For example, in Figuig Berber the cluster *lɖ* becomes the tense pharyngealized consonant *ʖ*, **tayəɖimt* gives *tayəʖlimt* ‘little ball of couscous’.

23.4.3.2. *The opposition lax versus tense*

In Berber morphology, consonants can alternate from lax to tense or vice versa. This is the case, e.g., in the verbal system, when we compare aorist forms with intensive aorist forms, or, in stative verb paradigms, the aorist forms with preterite forms. In the nominal system lax versus tense alternation can be found when one compares singular nouns with their plurals. The phonetic realization of the opposition lax versus tense varies from dialect to dialect and from consonant to consonant. Lax consonants are always realized shorter than their tense counterparts.

For some consonants the opposition is expressed in a difference in length, for example:

iləm 'skin' (Figuig)

illəm 'he has spun'

In other cases, the difference in length is supplemented by other phonetic features. Voiced lax consonants may have voiceless tense counterparts, e.g.:

mədl-əγ 'I buried' (Figuig)

məʔtl-əγ 'I always bury'

ʕəbr-əγ 'I measured' (Figuig)

ʕəppr-əγ 'I always measure'

Fricative lax consonants may have affricate tense counterparts, e.g.:

ħəsb-əγ 'I counted' (Figuig)

ħəʔtsb-əγ 'I always count'

Lax approximants may have tense plosive counterparts:

rəwl-əγ 'I flee' (Figuig)

rəggʷl-əγ 'I always flee'

A spirantized lax consonant may have a plosive tense counterpart, e.g.:

βði-γ 'I began' (Riffian)

βəddi-γ 'I always begin'

A lax voiced dorso-velar fricative may have a tense voiceless uvular counterpart:

ad i-nəγ 'he will kill' (Figuig)

ad i-nəqq 'he will continuously kill'

Some languages use the length opposition almost exclusively. For example, Tashelhit has only three cases where the length opposition is supplemented by some other change, viz., *ɖ* : *ʈʈ*, *ɣ* : *qq*, and *w* : *ggʷ*. On the other hand, in some Berber languages only *n*, *m*, *l*, *r*, *ħ*, and *ʕ* express the lax : tense opposition exclusively by length.

There has been considerable discussion about the status of tense consonants. Generative linguists consider them to be geminates, i.e., clusters of

two identical consonants. Most French scholars consider them to be monophonemic consonants. This problem is related to the insertion of ə. On the one hand, it is rare for a tense consonant to become divided into two parts by ə insertion. On the other hand, tense consonants behave differently from lax ones concerning open syllables. While ə insertion is impossible before a lax consonant followed by ə or a plain vowel, there is no impediment when a tense consonant is followed by ə or a plain vowel, e.g., Figuig Berber:

iʃrɹz → *iʃrəz* 'he cultivated' (**iʃərəz* is impossible)

iʃrɹz → *iʃərrəz* 'he always cultivates'

The issue is further complicated by the existence of a small number of words where two identical consonants may be separated by ə, e.g., Figuig Berber *imləl* 'it is white' (as opposed to *iməll* 'he became bored with'). This form suggests an opposition between a sequence of two identical consonants and a monophonemic tense consonant. Another possibility, however, would be to analyse ə in *imləl* as a structural, i.e. underlying, ə.

23.5. Some dialectal developments

Two developments have affected the phonological systems of many Berber languages. The first of these developments is spirantization, the development of lax stops into fricatives. The second is the gradual confusion of *w* and *y* with *u* and *i*.

23.5.1. Spirantization

Spirantization is a common feature in the northern part of the Berberophone territory. It reaches its culminating points in Riffian and Kabyle. Spirantization implies the development of lax stops into fricatives, e.g., *b* becoming *β*. In some cases spirantization is accompanied by a change of place of articulation. Spirantized *d*, for example, is *ð*.

Spirantization never affects tense consonants. Moreover, it may be prevented in certain contexts. There is, for example, no dialect in which *t* in the cluster *nt* can be spirantized. Spirantized and unspirantized lax stops are in complementary distribution. In many cases subsequent laxification of some tense stops or borrowing has blurred this situation. This leads to a system with spirantized lax consonants, non-spirantized lax ones, and non-spirantized tense stops, each with phonemic status.

23.5.2. The development of *w* and *y*

An important phonological development is the confusion of *w* and *y* with *u* and *i* respectively. In some Berber languages, *w* and *y* can be opposed to *u* and *i* in any position, e.g., in Figuig Berber. In most languages, however, there is a neutralization of this opposition in certain environments. This tendency towards neutralization may lead to a situation in which *w* and *u* and *y* and *i* are virtually allophones of each other, as for example in Tashelhit.

23.6. Tashelhit

The consonants of Tashelhit are given in Table 23-1.

Table 23-1. Tashelhit Consonant Inventory

	lab	dent	dent phar	pal	vel	vel lab	uvu	uvu lab	phar	glot
stops		t	ṭ		k	k ^o	q	q ^o		
	b	d	ḍ		g	g ^o				
fricatives	f	s	ṣ	š	x	x ^o			ħ	
		z	ẓ	ẓ̣	ɣ	ɣ ^o			ʕ	
nasals	m	n								
trill		r	ṛ							
lateral		l	ḷ							
approximant				y		w				h

23.6.1. Lax versus tense

As stated in § 23.4, the intensification of a consonant coincides in many cases with length. In Tashelhit, this is valid for all consonants, except *w*, *ɖ*, and *ɣ*, where the process of intensification leads to a phonetically unexpected result:

- (1) lax *w* corresponding to tense *gg^o*:
aorist: *izwiɣ*, preterite: *zgg^oaɣ* ‘to be red’ (a stative verb)
singular: *adgg^oal*, plural: *idulan* ‘in-law’
- (2) lax *ɖ* corresponding to tense *tɖ*:
aorist: *fɖɖ*, intensive aorist: *fɖɖr* ‘to have breakfast’

(3) lax y corresponding to tense qq :aorist: yrs , intensive aorist: $qqrs$ 'to slaughter'aorist: $y\bar{r}$, intensive aorist: $aqqra$ 'to read'aorist: $imyur$, preterite: $mqqur$ 'to be great' (a stative verb)

Lax–tense differences can play a role in Tashelhit dialectology. One dialect may have a lax variant of a lexical item, where another dialect has a tense variant, e.g., 'flour' is agg^om in the Aštukn dialect corresponding to $awwrn$ in Igdmíwn.

23.6.2. Labialized consonants

Besides the phonemes k , g , x , y , and q one finds the labialized forms k^o , g^o , x^o , y^o , and q^o . There are no minimal pairs of labialized versus unlabialized phonemes. In general we observe that some forms are more frequent than others. One finds more frequently: ak^or 'to steal', but the unlabialized variant exists also. On the other hand, we would never find $*imik^or$ 'thief' or $*arg^oaz$ 'man', but only $imikr$ and $argaz$.

23.6.3. Vowels, syllabification, shwa-insertion

The vowels are i , a , and u . One finds $[ə]$ on the phonetic level; it is, however, not a phoneme. The main preoccupation of recent phonological studies is the search for syllable build-up. None of the existing proposals is fully convincing. When one starts to apply proposed syllable structures to actually existing words, one sees that they do not give the expected surface result. An interesting analysis of Tashelhit syllabification and $ə$ -insertion was presented by Dell and Elmedlaoui (1988) for the Imdlawn dialect, in which a simple system for syllabification is attached to a sonority scale for the units that form the nucleus of the syllable.

23.7. Riffian

23.7.1. Spirantization

Riffian includes many spirantized consonants. In most Riffian dialects all lax stops except q have become fricatives, e.g., in the dialect of the Beni Said tribe:

 $b \rightarrow \beta$ $d \rightarrow \delta$ $\bar{d} \rightarrow \bar{\delta}$ $t \rightarrow \theta$

$g \rightarrow y$ or j

$k \rightarrow \check{s}$ or $\check{c}$

In other dialects, g has become $\check{z}$, while in some western variants g and k are stops. Moreover, in a number of dialects b has remained a stop.

Spirantization occurs everywhere, except when the consonant is preceded by a homorganic nasal:

$\theta\beta amb\check{s}\theta$ 'darkness'

$\theta anda$ 'pool'

$n\check{d}u$ 'jump!'

$antun$ 'yeast'

In these four examples, spirantization does not occur in the clusters mb , nd , $n\check{d}$, and nt .

There are some cases of lax stops outside this environment. They are either recent loans (e.g., $dija\check{z}$ 'loosen' from French *dégager* with d instead of $\check{d}$) or the result of a change of tense consonants to lax consonants. The precise conditions under which some tense stops have become lax stops are not clear.

The introduction of phonemic non-spirantized stops has lead to a tripartite system. First, we find a series of spirantized consonants. Second, we find the marginal series of non-spirantized stops. Third, we find the series of tense stops (which are never spirantized). On morphological grounds, it can be shown that the spirantized consonants are the lax counterparts of the tense stops, e.g.:

$ddaa$ 'live!'

$\theta u\check{d}aa\theta$ 'life'

Here, dd in the verbal form corresponds to $\check{d}$ in the noun.

23.7.2. *Shwa insertion*

The insertion of $\check{a}$ is conditioned by the structure of the word, according to the rules given in § 23.4. Some words do not follow the rules, and should be analyzed as containing phonemic $\check{a}$, e.g., $\check{z}\check{z}\check{a}h\check{d}$ 'strength' instead of the expected $*\check{z}\check{z}h\check{a}\check{d}$. Only in the case of r does sonority seem to play a role in the rules for $\check{a}$ -insertion. The development of r is complicated; it will be treated below.

23.7.3. *Particular developments*

Some phonemes have undergone developments which are typical for Rif Berber. Most important are the developments of l , ll , r and rr .

Except for some variants on the periphery of the Riffian territory, *l* has developed into an *r*-like sound, which is transcribed here as *ř*. Its phonetic realization varies from region to region. One may hear tap-like sounds, but also sounds similar to Czech *ř*. In most dialects *ř* is different from *r*, e.g.

θisiřa 'sandals' (← *θisila*)

θisira 'mills' (← *θisira*)

However, in some variants these two words are pronounced identically.

The tense counterpart of *l* has developed into the affricate *ddž*:

addži 'brains' (← *allī*)

ddžirəθ 'night' (← *lliləθ*, a loan from Moroccan Arabic *l-lila*).

In a number of loans *l* and *ll* occur:

lišanəš 'petrol' (← French *l'essence*)

βəlləʃ 'close!' (← Eastern Moroccan Arabic *bəlləʃ*)

However, in many loanwords *l* and *ll* have become *ř* and *ddž*.

The development of *r* and *rr* is more complicated. In Riffian, *r* is the only consonant which forces *ə* to be inserted before it. In most Riffian dialects, *r* and *rr* have had particular developments that can be summarized in stages as follows (we use V for *a*, *i*, or *u* and C for a consonant or a morpheme boundary):

stage 1:

<i>ər</i> → <i>ar</i>	<i>içərz</i> → <i>içarz</i> 'he plowed'
<i>ərr</i> → <i>arr</i>	<i>içərrəz</i> → <i>içarrəz</i> 'he plows'

stage 2:

<i>arC</i> → <i>aaC</i>	<i>içarz</i> → <i>içaaz</i> 'he plowed'
<i>arr</i> → <i>aař</i>	<i>içarrez</i> → <i>içaarəz</i> 'he plows'
<i>urC</i> → <i>oaC</i>	<i>θamurθ</i> → <i>θamoaθ</i> 'land'
<i>urr</i> → <i>oař</i>	<i>aħurri</i> → <i>aħoaři</i> 'free man'
<i>irC</i> → <i>eaC</i>	<i>aðβir</i> → <i>aðβea</i> 'pigeon'
<i>rV</i> → <i>rV</i>	<i>ari</i> → <i>ari</i> 'write!'

stage 3:

<i>aa</i> → <i>a</i>	<i>içaaz</i> → <i>içaz</i> 'he plowed'
<i>ř</i> → <i>r</i>	<i>içaarəz</i> → <i>içarəz</i> 'he plows'
	<i>aři</i> (← <i>ali</i>) → <i>ari</i> 'go up!'

In the first stage of the development, the opposition between *ə* and *a* is neutralized before *r*. In the second stage, *r* becomes a (rather short) *a*-like sound if not followed by a vowel. When preceded by *a*, this sound merges into a long vowel; when preceded by *u* or *i*, the diphthongs *oa* and *ea* emerge.

These two diphthongs may eventually develop into *wa* and *ya*. In the third stage of the development, the long *a* is shortened and at the same time *ř* (as found in the reflexes of *ir* and of *l*) merges with *r*. The importance of the mergers in the last stage may be exemplified by the word *ira* 'play!'. From the form in the third stage of the development it is not possible to predict whether the original form was **irar*, **irər*, **ira*, **ilar*, **ilər*, or **ila*. If we look at the form of the word in the second stage, *iraa*, only **irar* and **irər* are possibilities. From evidence of dialects where *r* has not undergone any change, we know that the proto-form was **irar*.

The three stages described above are not only historical stages, but are actually represented in the different dialects. The original situation can be found in Beni Iznassen (East), the first stage in Gueznaia (South), the second stage in Ait Said (Center), and the third stage in Ait Sidhar (North).

The question is whether *r* remains as an underlying phoneme in the three stages mentioned above or not. The answer is largely a theoretical choice. It should be noted, however, that these developments have not affected the morphology of the dialects. There has not been any analogical restructuring as a result of the phonetic developments.

23.7.4. Assimilations

The most important assimilations are:

A voiced non-liquid consonant becomes voiceless when immediately followed by *θ*:

aʕɾaβ 'Arab man'

θaʕɾaθ 'Arab woman'

The consonant cluster *ř + θ* becomes *tʃ*:

ðwəř 'return!'

θaməðwətʃ 'return (noun)'

The consonant cluster *mð* becomes *nd*:

θandint 'town' (← Arabic *mdina*)

Two identical lax consonants merge into their tense counterpart when in immediate contact:

nyəθ θ → *nyətt* 'kill (pl.) him!'

The consonants *ð* or *d* followed by *θ* or *t* result in *tt*:

ð θandint → *ttandint* 'it's a town'

Similarly *ð* or *d* followed by *θ* or *t* result in *tt*:

θayruðθ → *θayruʔt* 'shoulder'

23.7.5. *w* and *y*

w and *y* are opposed to *u* and *i* in all environments except at the end of a word. Here *w* becomes *u* and *y* becomes *i*:

nəḏwəɣ 'I jumped' (stem: *nəḏw*)

yənɗu 'he jumped'

ulyəɣ 'I went up'

yuli 'he went up'

23.8. Touareg

The phonological system of Touareg is quite different from that of the other Berber languages. It should be noted that especially in the field of the Touareg vowel system there is great uncertainty about the status of the different elements. This uncertainty is partly caused by internal complications of the system, partly by the variation across the different dialects and, last but not least, partly by the fact that much of the discussion is based on data collected at the beginning of this century.

23.8.1. *Vowels*

There are at least six vowels: *i*, *e*, *ə*, *a*, *o*, *u*. The vowel *ə* occurs in open syllables and is phonemic. There has been considerable discussion about the existence of a second short central vowel, transcribed here as *ä*, in opposition to *ə*. Recently, the existence of this phoneme has been shown with the help of phonetic instruments in Abalagh Touareg (Niger) by Naima Louali, featuring the following minimal pairs (1990: 138):

ägru 'understand!'

əgru 'find!'

älu 'weep!'

alu 'be like!'

This distinction is, however, not common to all dialects.

23.8.2. *Vowel quantity*

Most scholars distinguish long and short vowels. All vowels except *ə* and *ä* have a long variant. Length is said to play a role especially in the verbal system, where a number of tenses are distinguished by length. The study by Louali of the Abalagh Touareg system does not find any difference between short and long vowels. According to her, the tense differences which are said to be expressed by vowel length are expressed in reality by qualitative dis-

tinctions. Whether this situation pertains only to Abalagh or also for other dialects remains to be studied.

23.8.3. Consonants

In general, the Touareg consonant system is less rich than the systems of the northern dialects, due to the fact that Arabic has had less impact on Touareg and that most Arabic loans have been phonetically integrated into the Touareg system. Therefore, consonants like *h*, *ʕ*, and *ʂ* and their tense counterparts are absent. For the same reason, simple *q* and *t* are rare, though *t* may be the result of assimilation of *ɗ* to a following voiceless consonant. Cf. Ayr *eɗəs* 'sleep' corresponds to Iwlemmeden *eʔəs* 'sleep'. Resulting from the influence of Songhay and Hausa, some southern Touareg dialects feature *ŋ* as a borrowed phoneme. Special mention is warranted for two consonants: *z* and *h*. In Touareg, simple **z* has undergone various developments. In the north (Ahaggar), it has become *h*. In Mali one finds *ʃ* and in Niger *ʒ*. This can be illustrated by the name of the language in the different dialects. In Algeria Touareg is called *tamaahaq*, in Mali *tamaʃəq*, in Niger *tamaaʒaq*. These terms correspond to *tamaziɣt* in other Berber languages. The tense *zz* is always retained. By analogical formation, simple *z* has been reintroduced in a number of lexical items.

The consonant *h* is frequent in Touareg. In the Algerian variants it may be the correspondent of **z*, but it also occurs in words without **z*. This second *h*, which is also found in the other dialects, corresponds generally with zero in the Berber languages of the north and *β* in Ghadamsi:

ar 'lion' (Ouargla, Zenatic)

āhār 'lion' (Ayr)

aβor 'lion' (Ghadamsi)

There are a number of instances of palatalization. The consonants *g* and *k* become palatalized *gʲ* and *kʲ* and eventually develop into *ǧ* and *č*. These may become *ʒ* and *ʃ*. Before *i*, *t* may become *ʃ* in dialects of Niger.

Selected Bibliography

There are some good bibliographies concerning Berber linguistics. For the period until 1954 one can use the bibliography contained in Basset 1969. For the period 1954–1977, see Galand 1979. For the period after 1977, see Chaker 1991. The most readily available bibliography for American readers is Applegate 1970. Other titles given in the bibliography below all explicitly concern phonological subjects.

- Alojali, Ghoubeïd. 1980. *Lexique Touareg-Français*. Copenhagen: Akademisk Forlag.
- Applegate, Joseph R. 1970. "The Berber Languages." In *Current Trends in Linguistics*, ed. Thomas A. Sebeok, vol. 6: *Linguistics in South West Asia and North Africa*, pp. 586-661. The Hague: Mouton.
- Bader, Yusuf. 1985. "Schwa in Berber: A Non Linear Analysis." *Lingua* 67: 225-49.
- Basset, André. 1969. *La langue berbère*. London: Dawsons.
- Boukous, Ahmad. 1982. "Les contraintes de structure segmentale en berbère (dialecte tachelhit)." *Langues et Litteratures* 2: 9-27.
- . 1987. "Syllabe et syllabation en berbère." *Awal: Cahiers d'Etudes Berbères* 3: 67-82.
- . 1990. "Pharyngalisation et domaines prosodiques." *Etudes et Documents Berbères* 7: 68-91.
- Chaker, Salem. 1984. *Textes en linguistique berbère*. Paris: CNRS.
- . 1991. *Une décennie d'études berbères (1980-1990), bibliographie critique*. Algiers: Bouchene.
- Cortade, Jean-Marie. 1969. *Essai de grammaire touareg*. Algiers: Institut de Recherches Sahariennes.
- Dell, François, and Elmedlaoui, Muhammad. 1985. "Syllabic Consonants and Syllabification in Imdlawn Tashlhiyt Berber." *Journal of African Languages and Linguistics* 7: 105-25.
- . 1988. "Syllabic Consonants in Berber: Some New Evidence." *Journal of African Languages and Linguistics* 10: 105-25.
- Dell, François, and Tangi, Oufae. 1992. "Syllabification and Empty Nuclei in Ath Sidhar Rifian Berber." *Journal of African Languages and Linguistics* 13: 125-62.
- Galand, Lionel. 1953. "La phonétique en dialectologie berbère." *Orbis* 2: 225-33.
- . 1979. *Langue et littérature berbères, Vingt cinq ans d'études*. Paris: CNRS.
- . 1988. "Le berbère." In *Les langues dans le monde ancien et moderne*, ed. Jean Perrot, pt. 3, *Les langues chamito-sémitiques*, ed. David Cohen, pp. 207-42, 303-6. Paris: CNRS.
- Guerssel, Mohammed. 1983. "A Phonological Analysis of the Construct State in Berber." *Linguistic Analysis* 11: 309-30.
- . 1985. "The Role of Sonority in Berber Syllabification." *Awal: Cahiers d'Etudes Berbères* 1: 81-110.

- . 1986. "Glides in Berber and Syllabicity." *Linguistic Inquiry* 17: 1–12.
- Kossmann, Maarten G. forthcoming a. "Schwa en berbère."
- . forthcoming b. *Grammaire du Berbère de Figuig (Maroc Oriental)*.
- Lanfry, Jacques. 1968. *Ghadamès, étude linguistique et ethnographique*. Fort-National, Algeria: Fichier de Documentation Berbère.
- Leguil, Alphonse. 1981. "Remarques sur la labio-vélarité en berbère." *Bulletin de la Société de Linguistique de Paris* 76: 20–23.
- . 1982. "La phonologie au secours de la grammaire en Touareg." *Bulletin de la Société de Linguistique de Paris* 77: 341–63.
- . 1983. "La corrélation de concomitance en touareg." *Bulletin des Etudes Africaines de l'INALCO* 6: 77–123.
- Louali, Naïma. 1990. "L'emphase en berbère, étude phonétique, phonologique et comparative." Thèse pour le Doctorat en Sciences du Langage, Université Lumière Lyon 2.
- Petites Sœurs de Jésus, Agadez, Niger. 1968. *Initiation à la langue des Touaregs de l'Air*. Paris: Société d'Etudes Linguistiques et Anthropologiques de France (SELAF).
- . 1974. *Contes Touaregs de l'Air* (with a grammatical introduction by L. Galand). Paris: Publié avec le concours du CNRS.
- Prasse, Karl G. 1972–74. *Manuel de grammaire touarègue (tahaggart)*, 3 vols. Copenhagen: University of Copenhagen.
- . 1984. "The Origin of the Vowels o and e in Twareg and Ghadamsi." In *Current Progress in Afro-Asiatic Linguistics (Proceedings of the Third International Hamito-Semitic Congress, March 1978)*, pp. 317–26. Amsterdam: Benjamins.
- Saib, Jilali. 1976. "Schwa Insertion in Berber: Un Problème de Choix." *Afroasiatic Linguistics* 3/4: 71–83.
- . 1978. "Segment Organization and the Syllable in Tamazight Berber." In *Syllables and Segments*, ed. Alan Bell and Joan Bybee Hooper, pp. 93–103. Amsterdam: North-Holland.
- Stroomer, Harry J. forthcoming. *Dictionnaire Tachelhit-Français*.
- Willms, Alfred. 1972. *Grammatik der südlichen Berberdialekte (Südmarokko)*. Glückstadt: Augustin.

This page intentionally left blank.

Awngi Phonology

Robert Hetzron

University of California, Santa Barbara

24.1. Introduction

Awngi is the southernmost representative of Central Cushitic, or Agaw, the language cluster that is the most important substratum of Ethiopian Semitic and has survived in scattered pockets in Ethiopia and Eritrea. It is spoken today in the Province of Agawmeder, Goğğam Governorate General in Ethiopia. When I visited the location in 1965–66, I was told that the number of speakers may be around 50,000. Unfortunately, this figure is completely unreliable. It is only an estimate given by local tax-collectors.

I had the opportunity to talk briefly with Awngi speakers of other areas within the same province and detected dialectal differences. This contact was too brief to yield any publishable information.

In general presentations of Ethiopian languages, one often sees reference to an Agaw “language.” This is decidedly incorrect. The difference between the various Agaw tongues—the northern Bilin, the eastern Xamir, the western Kemant, and the southern Awngi—is greater than the diversity of the Romance languages.

In the older literature, Awngi was referred to as “Awiya,” a term introduced by Conti Rossini (1905). This important Italian scholar apparently had relatively little time to work on Awngi, and did not realize that the term *awíya* actually means ‘Agaw person’, lit., ‘son of Agaw’, fem. *awíja*, pl. *awayírí*. I introduced the term “Southern Agaw” (Hetzron 1969), but my colleagues preferred the name used by the speakers themselves, *awŋi*, with the derivative *-ŋi* used for names of languages (equivalent to, and indubitably the etymon of, Amharic *-ñña*). The *aw(V)* portion of these words was demonstrably **aɣʷ-* in the older language, the source of the Amharic name *agāw*.

My fieldwork¹ on Awngi was carried out in 1965–66 in Ethiopia, mainly in Addis Ababa, with a field trip to and a stint of ten days at the home of my

1. Sponsored by the Ford Foundation via the Near Eastern Center and the African Studies Center of the University of California, Los Angeles.

main informant, Mr. Asmare Tegenye, then a 22-year-old student, from the village of Bərcí (Amharic *Bərṭa Abbo*), subdistrict of Fafa, district of Ankäša, province of Agawmeder.

The oldest Southern Agaw document comes from the 1770s, a translation from Amharic of the Song of Solomon (Ms. 33, Ethiopian Collection, Bodleian Library, Oxford, “Damot Agaw,” fols. 18–20), commissioned by the Scottish traveler James Bruce. This text is still unpublished. It is of mediocre quality. Basically, the translator applied a word-by-word translation from an Amharic version (which, also commissioned by Bruce, was translated from Ge‘ez) without much adjustment for style and grammatical continuity. He used the Amharic script (with inconsistencies), yet it is valuable for purposes of historical morphology.

24.2. The phonemes

24.2.1. *The consonants*

Length is in principle relevant for consonants, but geminate consonants are rare. Minimal pairs are few, either in some Amharic borrowings: *fetenəŋ* ‘do quickly’ (Amh. $\sqrt{f}t\dot{n}^2$ vs. *fettenəŋ* ‘give an exam’ (Amh. $\sqrt{f}t\dot{n}$), or involve morpheme boundaries: *dadá* ‘path’ (diminutive-feminine of *dad* ‘road’) vs. *daddá* ‘on the road’. Yet geminates occur in genuine Awngi words as well: *əttiŋ* ‘to fall’, *əttińí* ‘flour’ (the “fall-out” of milling), *cəllí* ‘little’, *kuppi* ‘fruit’, *angučča* ‘cat’, etc.

Table 24-1. Awngi Consonants

p	f	t	c	s	š	č	k	k ^w	q	q ^w
b		d	ɟ	z	ž		g	g ^w	ɣ	ɣ ^w
m		n					ŋ	ŋ ^w		
w					y					

The symbols *c/ɟ* stand for the affricates *ts/dz*.

As in Amharic, *b* is [β] in intervocalic and word-final positions. There is no *v*. The *d* is slightly retroflex. Even though Awngi has *ɟ* (which only occurs in the numeral *seza* ‘4’ and its derivatives), in morphophonemic alternations

2. This may be a recent borrowing. Amharic *t* is usually rendered by *c* in Awngi, e.g., *kecer* ‘make an appointment’, Amh. $\sqrt{qt}$. In the following, glottalization is marked by an underdot, except for *ɟ*, which stands for a retroflex.

the voiced counterpart of *c* is *z* (most certainly reflecting older articulations).³ Unless geminated, *r* is realized with only one flap. It does not occur in absolute initial position. Both *q* and *ɣ* are lax, but not fricative. Labio-velars and labio-uvulars are distinct from the homologous velars and uvulars only before *-a*, *-e*, *-i* and word-finally. In morphophonemic alternations, both semivowels, *w* and *y*, assume the status of the voiced member of the pair.

24.2.1.1. Some consonantal changes

Not less surprising than the etymological relatedness of French [fis] and Spanish [ixo] 'son' is the fact that Awngi *ɲárí* and Somali *madaḥ* (< **math*), both meaning 'head', are perfect cognates. (The final *í* of Awngi represents the palatalization caused by the final pharyngeal). Two historical changes explain this.

I wish to point out that at the present stage of research, hardly any work has been done on historical phonology of Agaw. The contexts in which the changes took places, as against those in which the phoneme kept its phonetic character, are unclear.

24.2.1.1.1. *m* > *ɣ*

Awngi *ɲ* often corresponds to *m* in other Cushitic languages, but to *ɲ* or *n* elsewhere in Agaw. Unlikely as it looks, surmising a *m* > *ɲ* change for Proto-Agaw is inevitable.⁴

The root *ɣu-* 'eat' corresponds to **kom* in East Cushitic, suggesting that in some positions different types of changes took place as well.

The Agaw languages also have /*m*/. The only clear attestation of a *m* that comes from Afroasiatic is in the question particle *-mà*. For the rest, most of them are obvious borrowings from Ethiopian Semitic. A number of them cannot be accounted for, and the only significant statement is that they seem to have no Cushitic etymology.

24.2.1.1.2. **t* > *r* > *y*

Let me illustrate the changes through the borrowing from Semitic (')*amät* 'year' which follows the behavior of native words. Bilin regularly has *r* where Proto-Cushitic **t* must be posited: *amära*. On the other hand, *-r* → *-t* change is one of the plural-forming devices: *šəmar* → *šəmat* 'tail' (Palmer

3. In Palmer's 1957 material, there are several occurrences of *ɜ* where my informants had *z*, a clear case of age-dialect (his informant was from the same area as mine).

4. Diakonoff's (1992: 33) suggestion that *ɲ* arose through merger with the next phoneme is untenable.

1958: 385, cf. Xamir *šəmər*, Awngi *cəmār*),⁵ though ‘year’ keeps its *r*. Kemant has *y*: *amäy*. As in Bilin, one of the nominal plural-forming devices is *r* → *t*, and it also applies to *ámra* → *amət* or *émət* ‘year’ (Appleyard 1984: 41, 1987: 255). Awngi has *amet* (which may also be a recent borrowing), but the overall situation is more complicated.

Albeit rarely, *r* < **t* is indeed found in Awngi, as in *-rŋa*, the ending that makes ‘20’ out of ‘2’ (elsewhere it is the 10-formative in general: Bilin *-rāŋan*, Kemant *-yəŋ*), coming from Cushitic **tVmVn* ‘ten’. The verb for ‘die’ has the altering stems *kət/kə/kər-* where the final consonant may ultimately represent the “autobenefactive” ending, Cushitic **d*.

More frequently, **t* may either remain *t* or become *y*, as in Kemant. One important case in point is the onset of the Sg./Pl.2/Sg.3f. ending of the verbs that comes from Proto-Cushitic/Afroasiatic **t-* (the suffixes come from prefix-conjugated auxiliaries). Bilin has *-r-* for Sg.2/3f. and *-d-* for Pl.2, Kemant only *-y-*.⁶ Xamir has the alternation *-r- ~ -d-* or *-dr-(?)* for the entire domain of this person-marker, *-d-* after *l*, *n*, *r*, and *-r-* everywhere else (Appleyard 1987: 478). Appleyard (1987: 472–73, 476) also mentions the alternation *-t- ~ -r-* for the autobenefactive ending: Pl.1/3 and the indicative Sg.3f. exhibit *-r-*, the rest *-t-*, including subordinate Sg.3f.

Let me first mention the root *bay-* ‘leave’, cf. Bilin *bār-*, Xamit *bār-*, Kemant *be-*. This type of correspondence suggests an original **bat-*, which does show up in the conjugation as *bay-* as well.⁷ Moreover, the person-marker **t* does have palatal manifestations in Awngi (§ 24.4.3). Let us consider the following comparative table of some persons of the Indefinite Nonpast in three verb classes where classification is based on the choice of suffixes:⁸

5. Reinisch (1887: 325) compares this word with Hebrew *šmr* ‘bristle, stiffen’, particularly *sāmār* ‘wiry-haired’ and Aramaic *mr* ‘wool’ (also *qmr*) and to Ge‘ez *dämr* ‘wool’ (Hebrew also has *šmr* for ‘wool’). It does not seem to be a borrowing, but rather common Afroasiatic heritage, starting with the phoneme that Diakonoff (1992: 15–16) reconstructs as **š*. The Aramaic root does not fit into the picture, since it has what Diakonoff (1992: 20–21) reconstructed as **č*. Ge‘ez also has the verbal root *ḏmr* for ‘attach’, so that the semantic feature connecting all of this may be “appendage.”

6. Archaic *-t-* is still used in four verbs, all of which ending in *-y*, which blocked the change *t* > *y* (Appleyard 1975: 332).

7. Appleyard (1987: 473 n. 38) compares this to East Cushitic **bah* ‘go out’ + autobenefactive **d*, a retroflex.

8. Only the first hyphens are to be taken as clear morpheme boundaries. The later hyphens are added to facilitate the explanation; cf. note 9. There are two more verb classes which do indeed involve *t*’s and *y*’s (the latter even in the Sg.3m.) and combinations of both (second

		A	B	C
I:	Sg.1	<i>des-é</i>	<i>qúc-é</i>	<i>bá-t-é</i>
II:	Sg.2=Sg.3f.	<i>des-té</i>	<i>qúc-í</i>	<i>bá-t-é</i>
III:	Sg.3m.	<i>des-é</i>	<i>qúc-é</i>	<i>báy-é</i>
IV:	Pl.1	<i>des-né</i> 'study'	<i>qúc-né</i> 'wash'	<i>bá-né</i> 'leave'

A glance at this comparative table suggests the following. Class A exhibits the relatively most archaic features. Class B uses a II form where the person-marker $t > y$. Class C must have had a root ending in $*-t$ which became $-y$, except in I.⁹ The $-y$ -disappeared in IV.

24.2.1.1.3. THE SG.3F. INDEPENDENT PRONOUN

Here is a manifestation of $t > y$ that has typological interest. The fact that Awngi, a language that has grammatical (though semantically determined) gender, does not have it in its Sg.3 pronoun, is in contradiction to one of Greenberg's (1966: 96) universals (43): "If a language has gender categories in the noun, it has gender categories in the pronoun." The explanation is historical. On the basis of comparison with other Agaw languages one may reconstruct $*\eta i$ for 'he' and $*\eta it$ for 'she'. The latter underwent a series of changes: $*\eta it > *\eta ir > *\eta iy (> *\eta i) > \eta i$, each stage of which, other than the starting point, is attested in various Agaw languages (Bilin *nirí*, Xamir *ŋír* or *ŋí*, Kemant *niy*). Thus, in the last reconstructed stage there may have been an opposition $*\eta i/\eta i$: or $*\eta iy$ for 'he/she', and with the loss of length or reduction of the diphthong, the two collapsed.

24.2.2. The vowels

Vocalic length is not relevant in Awngi. Table 24.2 presents the vowels.

persons+Sg.3f.) for which I have no explanation except for the assumption that the palatal element may have been part of the root.

It should be mentioned that proto-Cushitic may have had $*d$, $*t$, glottalized $*d'$, $*t'$, and retroflex $*\bar{d}$, as against a poorer system in Agaw. Our $*t > r > y$ may come from $*t$, as in the person marker, and from $*\bar{d}$ as in the root final assimilated autobenefactive. In Xamir, however, there is $r \sim t$ for the autobenefactive, but $r \sim d/\bar{d}r$ for the person marker. In both Bilin and Xamir, the plural shows an alternation $r \sim t/\bar{t}$, whereas the person marker has a voiced consonant as the second member.

9. Palmer (1959: 283) considers this a voice-pairing y/t in the sense of my § 24.4.2. Two objections may be raised. Historically, the first members of the pair are more archaic in § 24.4.2, whereas here t is the relic. Descriptively, one ought to set up a y -deletion rule for II and IV (the clusters $-yt-$, $-yn-$ are otherwise allowed), and the situation in C would be even more complicated.

Table 24-2. Awngi Vowels

i		[ə]	u
	e		o ~ wa
	æ		
		a	

REMARKS: *æ* occurs only at a morpheme boundary, the result of a contraction of a preceding palatal element and a subsequent *a*, thus *ŋi-* ‘his/her/its’ + *-ará* ‘husband’ yields *ŋærá* ‘her husband’. Most occurrences of [ə] are predictable on the basis of syllabic and phonetic structure: the avoidance of triple clusters within the word, of double clusters word-initially, and of word-initial *ɿ* but they are relevant in the tonal rules (which are then ordered after the shwa-introduction), e.g., *śrí* ‘rain’. Yet, there are a handful of words beginning with *ə*CV-: the verbs *əŋəŋ* ‘bite’, *əq^uíŋ* ‘call’, *əśíŋ* ‘spend the day’, *əyəŋ* ‘give’, *əyayerəŋ* ‘curse’, and *əyoyíŋ* ‘laugh’, where the post-*ə* consonants have no positional constraints, cf., *yiməŋ* ‘beg’, *yasəŋ* ‘have on/with oneself’, etc.

The close vowels *i*, *ə*, *u* sound somewhat more open before uvulars and labio-uvulars. In a sequence /iʏ/, one hears something like [i^aʏ], comparable to the phenomenon of furtive *pataḥ* in Hebrew.

Amharic *ä* is rendered by the palatal vowel *e* in loanwords.

24.3. The tone system

The fact that Awngi is a tone language was first recognized by Palmer (1959). More and more tonal systems have been found in the Cushitic area, which inevitably leads to the conclusion that Proto-Cushitic must have been tonal. The northern Agaw language Bilin was described by Palmer (1960: 111) noncommittally as to tone. First of all, whatever the descriptive device, historically speaking in languages where a major word may have no stress at all (like Japanese and Bilin), one must suspect an earlier tonal stage where the now stressless word had no High tone. In Bilin the stressed syllable, with a high pitch, is followed by syllables carrying the same pitch, not unlike the case in Awngi (§ 24.3.4); but in absolute final position, the high tone is not realized. Thus the distinction between the minimal pairs *bəta* ‘dust’ and *bətá* ‘louse’ is perceivable only if they are followed by some other word (Palmer 1960: 113). In Awngi, indeed, tones play a greater role than in Bilin.

24.3.1. *The tonemes*

Awngi has four tonemes arranged on three levels: High *á*, Mid *a* (unmarked in this notation), Low *à*, and Falling *â* (a fall from High to Mid). The last two are quite rare. Low tone *à* (only with this vowel) occurs only in the past tense ending *-y"à* and the sentence-question particle *-mà*.¹⁰ The Falling tone always occurs in word-final, but not necessarily absolute final, position. Some minimal oppositions: *aqê* ‘man (accusative)’, *áqé* ‘I am/know’, *aqé* ‘don’t be!’; *kur* ‘hill’, *kúr* ‘saddle’; *yuna* ‘woman’, *yúna* ‘we have eaten’, *yúnà* ‘they ate’.

Since Awngi has no expressive intonation or downdrift to speak of, from a distance Awngi speech sounds like a sort of “twittering,” a succession of equitonal High and Mid pitches, with occasional Low tones. Tones may be recognized even out of context, when a single word is pronounced. Mid represents the normal “tessitura,” or level of voice, High is shriller, more elevated, though not falsetto, and a Low is creaky.

24.3.2. *Tonemic rules*

Before an added syllable, the Falling tone becomes High and the Low one Mid: *desûs* ‘while he studies’ has a variant *desúsi*, and it is *desúskí* before the multifunctional (connective-topical) particle *-kí*. The past tense form *desy"à* ‘he studied’ becomes *desy"amà* in interrogation: ‘Did he study?’.

24.3.3. *Stress*

Somewhat surprisingly, Awngi has some stress phenomena that operate independently of the tone system. Accompanied by a slight rise of the pitch, it normally falls on the penultimate syllable.¹¹ Yet it does play a role in two situations. Unlike other suffixes, the referential article *-ká* and the connective-topic marker *-kí* do not change the position of the stress (while the plural ending *-ka/ká* does).

10. While the former is always sentence-final, the question particle may also be placed after a focus-element, if any, which keeps it original position. In the absence of focus, it is the sentence-final verb that will carry it.

11. For example, for the postposition *falengá* ‘after’, I first transcribed a Mid-High-High tone pattern. My informant was linguistically sophisticated enough by that time to understand it and to object, suggesting Mid-Mid-High. When I asked him to whistle the word (a technique that I learned from the late A. N. Tucker), he unmistakably produced a Mid-Mid-High melody. Further familiarization with the Awngi sound system under the guidance of Mr. Asmare later allowed me to hear that the rise on the middle syllable was slighter than of the final one, as an exponent of stress.

There is a small class of nouns where the stress is not penultimate in that it may not leave the stem even if it is followed by more than one syllable (see § 24.4.5.2), e.g., *g'uzg_i-u_i-des_k* 'from_k that of_{j, masc. gen.} the stomach_i' vs. "normal" *giz-'u-des* 'from that of time' (i.e., genitives combined with another case-marker). The first type with stress restriction is very rare. Its origin is likely to be somehow connected with the fact that these nouns only have Mid tone syllables.

24.3.4. *The origin of tonemes*

In roots, i.e. words without suffixes, there may be Mid-High sequences, but not High-Mid ones (cf. the beginnings of §§ 24.4.5 and 24.4.6.1). Thus, once High tone is assumed, it remains High for the rest of the root, as is probably the case in Bilin. This may have been the Proto-Agaw situation, though not the Proto-Cushitic one. Descriptively, however, this is no longer true in view of the complications involving suffixes.

While I am not sure of the origin of the Low tone in the sentence question particle (language universals would have rising in interrogation),¹² plausible justification for it may be advanced in the case of the past tense suffix. One may reconstruct **-uya* with a principle of equitonicity for the two vowels. When *u* was reduced to ^u, i.e. lost its syllabicity and capability of carrying tone, the closest the originally equitonal *a* could do was to become Low, which thus properly stands for lack of tone.

The Falling tone always occurs in word-final position and comes from the contraction of two syllables (with no intervening consonant, thus **-iá* → *ê*, still so in Bruce's unpublished document).

24.4. Morphoph/tonemic rules

24.4.1. *Regressive vowel height assimilation*

See (Palmer 1959: 273, Johnson 1972: 74–5). Given a word where the last syllable of a word has *e*, when a suffix with *i* is added, *e* → *i*. This assimilation spreads regressively in a chain onto immediately preceding syllables. If they also have *e*, it also becomes *i*. While only stem-final *e* can trigger this assimilation, farther back in the chain *o* → *u* also takes place. All other

12. As the native speaker of a language (Hungarian) that has High-Mid melody at the end of a sentence question, I cannot help suggesting that the Low tone in *-mà* reflects the possibility that this particle started out as a sort of tag-question marker, "...?, or what," which does go down.

vowels stop the spread. Thus, with the masculine suffix *-í*, the root **mole-qés*- ‘monastic’ (from Greek μοναχός via Amharic) yields *muliqísí* ‘monk’, whereas the feminine exhibits the original underlying vowels: *moleqésá* ‘nun’. By itself, *o* cannot start this process. The word *gorýí* ‘spade’ has no raising. In my material there is no sequence of two *o*’s or *o* → *u* preceded by underlying *e*.

A few words, mainly Amharic borrowings, such as *šererití* ‘spider’, but also the indigenous *əyáyérí* ‘curse’, are exempt from this rule.

24.4.2. The root-final voice alternation

In the verbal system, the following alternations of root-final consonants are observed:

$b \rightarrow p$, $w \rightarrow p$, $d \rightarrow t$, $z \rightarrow c$, $\check{z} \rightarrow \check{c}$, $g \rightarrow k$, $y \rightarrow k$, $g^w \rightarrow k^w$, $\gamma \rightarrow q$, $\gamma^w \rightarrow q^w$

The first members of the pair are the unmarked ones, as is illustrated by the fact that two different consonants lead to *p* or *k*. The second ones occur (i) in the imperative, (ii) before a Sg.1 suffix. Examples: *cép-é* ‘I do’ vs. *céw-é* ‘he does’ and *cíp* or *cipí* ‘do_{sg.}’. Naturally, roots may end in voiceless consonants, in which case no alternation takes place.

The following explanation may be appropriate for the alternation in question. The singular imperative is the only form of the verb where the root-final consonant is in absolute word-final position, actually in sentence-final position because of strict verb-final syntax of Awngi from Proto-Cushitic on. This may have led to devoicing, a phenomenon that is well attested elsewhere. The plural imperative (and the imperative of verbs ending in a vowel) and the prohibitive may have adopted the devoicing by solidarity. Now, Cushitic verbal suffixes are known to have been independent prefix-conjugated auxiliaries based on the general Afroasiatic pattern (as in Semitic): *t*- for second persons and Sg.3f., *y*- for the rest of the third persons and *n*- for Pl.1. For Sg.1 we find *a*- in Cushitic prefix-conjugations, and ‘V- in Semitic.¹³ If we take ‘V- to be the original prefix (in proto-Agaw at least), the addition of -‘V- to the verbal stem as part of the prefix of the auxiliary-to-become-suffix, this ‘created an environment comparable to absolute final position, hence the devoicing.

13. The glottal stop is not automatic, as shown in Classical Arabic where automatic glottal stops drop after a prefix (*wasla*), but not in Sg.1, e.g., *‘al-malik*- ‘the king’, but *wa-l-malik*- ‘and the king’ with the glottal stop dropped, vs. *‘asma*- ‘I hear’ and *wa-‘asma*- ‘and I hear’.

24.4.3. *The palatal -y- (< *t) suffix-person-marker (see § 24.2.1.1.2)*

In several verb classes, Sg.3f. and Sg./Pl.2 have a palatal person-marker¹⁴ followed by the rest of the suffix expressing tense, mood and/or type of subordination.¹⁵ Before a consonant, it manifests itself as *i* or *í*: *kantiy*^wà ‘she saw’. It palatalizes through merger with the subsequent vowel in the following manners. (i) with a nonfinal *-a-* it merges into *-e-*: **kanti-ání* → *kanténí* ‘when she sees/saw’. (ii) With a word-final *-a* it yields *-æ*: **kanti-a* → *kantæ* ‘she has seen’; yet when this becomes nonfinal through the addition of a further suffix *-æ-* → *-i-*: **kantæ-štíy*^wà → *kantištiy*^wà ‘she was seeing’. (iii) It blends with a subsequent *-u*, *-e*, *-ə* (and vacuously with an *-i*) into *-i-*: **kanti-ûs* → *kantís* ‘while she is/was seeing’ (note the change of tone).

24.4.4. *Tonal dissimilation and assimilation*

Two case-suffixes, the locative *-da* and the ablative *-des*, have the tone opposite to that of the preceding syllable: Mid after High and High after Mid, e.g. *murí-des* ‘from (a/the) village’ vs. *múri-dés* ‘from (a/the) snake’. Yet the tone is always Mid when the preceding syllable is a genitive ending (in Awngi, genitives agree in sentence-case with their headnouns, see Hetzron 1994, and § 24.4.5.2 for an example).

The adverbial ending *-ŋa* ‘in the manner of, like’, on the other hand, always has the same tone as the preceding syllable (final vowels of stems drop): *mur-ŋa* ‘like a village’ and *múr-ŋá* ‘like a snake’.

The other case-endings have their own tones.

24.4.5. *Tonal phenomena in nominals*

Nominals possess a lexically determined tone pattern which is constant except for the few illustrated in § 24.4.5.1. High-Mid sequences within a stem are very rare; I only found a reduplicative *bádbaday* ‘dove’, *éyay-i* ‘ice’, and *cémark*^w-í ‘eyelash’.¹⁶

Five parameters are to be used to determine the nominal classes (also including adjectives). Two of them apply only to a handful of nouns.

14. Other verb-classes still have the original *-t-*. Second persons and Sg.3f. having the same morphological marking is a feature of Afroasiatic prefix-conjugations. The Cushitic conjugational suffixes come from original prefix-conjugated auxiliaries.

15. In Awngi suffix-conjugations the person-marker never stands alone.

16. Could this be a compound (cf. *ərk*^w-í ‘tooth’)? The interrogative pronoun *wóšíní* ‘which’ is clearly composite: with an adjectival derivative *-in-í*.

24.4.5.1. *Tone Lowering*

Most nominals have a consistent tone pattern, the same all through the inflections. However, a few have a Mid-(Mid-)High tone pattern with the last High lowered before the plural ending or before the adverbial case-suffix (the first one is an Amharic borrowing, the third comes from Arabic):

agér ‘country’, *ager-ká* ‘countries’, *ager-ŋa* ‘as a country’

areméc ‘embers’, pl.: *aremec-ká*

fírísí ‘horse’, pl.: *feres-ka*

Note that the last noun has Mid tone on the plural ending.

24.4.5.2. *Stem stress (see end of § 24.3.3)*

For a few nominals that have only Mid tones, High plural ending and Mid accusative suffix, the stress may not leave the stem: *g’uzg-u-des* ‘from that^{masc.} of the stomach’, as against *gurgóm-u-des* ‘from that of the neck’, which exhibits the majority behavior.

24.4.5.3. *Post-stem vowel*

The feminine ending may be *-a* or *á*, according to the class. The corresponding masculine may end in *-i/-í* with the same tone or in a consonant.

24.4.5.4. *Plural*

The plural ending may be *-ka* or *-ká*, a further criterion for classification (cf. § 24.4.5.1).

24.4.5.5. *Accusative*

The accusative ending is *-o/-ô* after consonant, where the tone is as a classificatory factor. After masculine *-i/-í*, it merges with it yielding *-e/-é*, after another vowel it is *-wa*. These latter are thus not relevant for classification.

24.4.6. *Tone in verbs*

24.4.6.1. *Tone patterns of stems*

With only a few exceptions, no matter what the tone-pattern of the verb stem is, infinitives only have Mid tones. They end in *-əŋ*, most probably from Afroasiatic **m-*.

As for the finite forms, most verbs have a constant tone-pattern for all the conjugations (I Mid and II High below). However, there are a few monosyllabic verbs, mainly prefix-conjugated ones and one suffix-conjugated (*cewŋ*) that have High tone in the Nonpast and forms derived therefrom, and a Mid

one in the Past and its derivatives (III and IV below both have the Raising feature (§ 24.4.6.2), hence the *-ú-*). Moreover, a handful of verbs¹⁷ have different tone-patterns according to the person within the conjugation (group V below):

Sg.3m.	I	II	III	IV
Indef.Nonpast	<i>des-é</i>	<i>tás-é</i>	<i>céw-é</i>	<i>y-ínt-é</i>
Def.Past	<i>des-ɣʷà</i> 'study'	<i>tás-ɣʷà</i> 'hit'	<i>cew-úyà</i> 'do'	<i>y-int-úyà</i> 'come'
but	V			
	Indef. Nonpast	Imperative		
Sg.1/3m.	<i>gemé</i>	sg. <i>gém</i>		
Sg.2/3f.	<i>gembé</i>	pl. <i>gemán</i>		
Pl.1	<i>gémné</i>			
	'go down'			

24.4.6.2. Tonal influence on suffixes

Many verbs have the quirk of exerting a tonal influence on the first syllable of the subsequent suffix, doubtless the remnant of a dropped stem-final syllable, probably a vowel (a connective vowel in Bilin). Indeed, when there is no suffix, in the Sg.m. Imperative, a final *-i* carrying the tone projected from the root is added. When that syllable has the vowel required by the root anyhow, the rule applies vacuously. This allows a fivefold classification of the verbal roots:

(I) **Neutral**: roots that exert no influence on the subsequent syllable and thus allow the establishment of the basic tone of the suffixes.

(II) **Raising**: roots that require a Falling tone on the subsequent syllable if it is final (#), otherwise it will be High.

In the rest of the verb classes, the second persons and Sg.3f. are marked by a palatal element (§ 24.4.3):

(III) **Lowering**: roots that, except for Pl.1, require a Mid tone on the subsequent syllable. This is all monosyllabic verbs or reduplicative bisyllabic ones.

(IV) **3-Lowering**: roots that require a Mid tone on the subsequent syllable in the second persons and Sg.3f. only. Along with (I) and (II), this is a class with wide membership.

17. There are seven of them, all monosyllabic, four end in *-ŋ*, two in *-m*, and one in *-y* (*əyəŋ* 'give') where etymology would suggest a final **-w* [Kemant *iw-*, Xamir *yəw-*].

(V) **Raising-Lowering**: a root that require Mid tone on the subsequent syllable in the second persons and Sg.3f., and a Falling#/High one in all other persons. One verb only.

(VI) **Lowering-Raising**: a root where the Sg.3m./Pl.3 marker is palatal with Lowering, all the rest have Raising. One verb only.

(VII) **Partial Raising**: Raising in Sg.1 and in the second persons and Sg.3f. One verb only.

(VIII) **Partial Lowering**: Lowering in all second and third persons, but not in the first persons. Four verbs.

	I	II (R)	III (L)	IV (y-L)
Past Sg.3m.	<i>des-ɣ^wà</i>	<i>cew-úɣà</i>	<i>zum-uyà</i>	<i>kantɣ^wà</i>
Sg.3f.	<i>des-tɔɣ^wà</i>	<i>cew-túɣà</i>	<i>zum-iɣ^wà</i>	<i>kant-iɣ^wà</i>
Nonpast Sg.3m.	<i>des-é</i>	<i>céw-é</i>	<i>zum-e</i>	<i>kant-é</i>
Sg.3f.	<i>des-té</i>	<i>céw-té</i>	<i>zum-i</i>	<i>kant-i</i>
Pl.1	<i>des-né</i> 'study'	<i>céw-né</i> 'do'	<i>zum-né</i> 'talk'	<i>kant-né</i> 'see'
	V (RL)	VI (RL)	VII (PR)	VIII (PL)
Past Sg.3m.	<i>bəft-úɣà</i>	<i>əs-iɣ^wà</i>	<i>kəmbabl-iɣ^wà</i>	<i>fəč-iɣà</i>
Sg.3f.	<i>bift-iyà</i>	<i>əst-úɣà</i>	<i>kəmbabəl-túɣà</i>	<i>fəčt-iyà</i>
Nonpast Sg.3m.	<i>bəft-é</i>	<i>əs-i</i>	<i>kəmbabl-i</i>	<i>fəč-i</i>
Sg.3f.	<i>bift-i</i>	<i>əst-é</i>	<i>kəmbabəl-tí</i>	<i>fəčt-i</i>
Pl.1	<i>bəftə-né</i> 'be hidden'	<i>əsn-é</i> 'weep'	<i>kəmbabəl-né</i> 'roll about'	<i>fəč-né</i> 'hate'

As we can see, Pl.1 is never affected by Lowering; in fact it is never affected at all. Not all instances of palatal second person/Sg.3f. marker have Lowering, but there is no Lowering without at least one palatal element at the end of the root. When this one is in the Sg.3m./Pl.3 form, there is always Lowering. Thus, as in the case of the Low tone (§ 24.3.3), Lowering must have something to do with the palatal element **i* losing its segmental character in several contexts. As for Raising, it has its parallel in Bilin, in the verb class where the connective vowel carries the stress (Palmer 1957: 138 [ii]).

The causative verb *yíg^wcəŋ* 'save (for later)', the causative of the prefix-verb *yiguŋ* 'remain' represents a mixed type: Mid tone and Raising-Lowering for the past, and High tone with neutral suffix *s* for the Nonpast. Finally, Awngi has 13 short roots consisting of one consonant (in half the cases with a labial appendix *w*). They are idiosyncratic and need no analysis in a general study.

Bibliography

- Appleyard, D. L. 1975. "A Descriptive Outline of Kemant." *Bulletin of the School of Oriental and African Studies* 38: 316–50.
- . 1984. "The Internal Classification of the Agaw Languages: A Comparative and Historical Phonology." In *Current Progress in Afro-Asiatic Linguistics: Proceedings of the Third International Hamito-Semitic Congress*, ed. James Bynon (Current Issues in Linguistic Theory 28). Amsterdam: Benjamins.
- . 1987. "A Grammatical Sketch of Khamtanga, I–II." *Bulletin of the School of Oriental and African Studies* 50: 241–66, 470–507.
- . 1992. "Vocalic Ablaut and Aspect Marking in the Verb in Agaw." *Journal of Afroasiatic Languages* 3: 126–50.
- Conti Rossini, Carlo. 1905. "Noti sugli agaw, II. Appunti sulla lingua Awiya del Danghela." *Giornale della Società Asiatica Italiana* 18: 103–94.
- . 1912. *La langue des Kemant en Abyssinie* (Sprachkommission der Kaiserl. Akademie der Wissenschaften 4). Vienna: Alfred Hölder.
- Dolgopolski, Aron B. 1973. *Сравительно-Историческая Фонетика Кушитских Языков*. Moscow: Nauka.
- Diakonoff, Igor M. 1992. "Proto-Afrasian and Old Akkadian: A Study in Historical Phonetics." *Journal of Afroasiatic Languages* 4: 1–133.
- Greenberg, Joseph H. 1966. "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In *Universals of Language*, ed. J. H. Greenberg, pp. 73–113. Cambridge: MIT Press.
- Hetzron, Robert. 1969. *The Verbal System of Southern Agaw* (University of California Publications, Near Eastern Studies 12). Berkeley and Los Angeles: University of California Press.
- . 1976. "The Agaw Languages." *Afroasiatic Linguistics* 3/3: 31–75.
- . 1978. "The Nominal System of Awngi (Southern Agaw)." *Bulletin of the School of Oriental and African Studies* 41: 121–41.
- . 1994. "A Note on Genitival Agreement in Awngi." In *Double Case*, ed. Frans Plank. New York: Oxford University Press.
- Johnson, C. Douglas. 1972. *Formal Aspects of Phonological Description* (Monographs on Linguistic Analysis 3). The Hague: Mouton.
- Palmer, F. R. 1957. "The Verb in Bilin." *Bulletin of the School of Oriental and African Studies* 19: 131–59.
- . 1958. "The Noun in Bilin." *Bulletin of the School of Oriental and African Studies* 21: 376–91.

- . 1959. "The Verb Classes of Agaw (Awiya)." *Mitteilungen des Instituts für Orientforschung* 7: 270–97.
- . 1960. "An Outline of Bilin Phonology." *Atti del Convegno Internazionale de Studi Etiopici, Roma, 2–4 Aprile 1959*, pp. 109–15 (Problemi attuali di scienza e di cultura 48). Rome: Accademia Nazionale dei Lincei.
- Reinisch, Leo. 1887. *Die Bilin-Sprache*, II. *Wörterbuch der Bilin-Sprache*. Vienna: Alfred Hölder.
- Sasse, Hans-Jürgen. 1979. "The Consonant Phonemes of Proto-East Cushitic (PEC): A first approximation." *Afroasiatic Linguistics* 7/1: 1–67.

This page intentionally left blank.

Oromo Phonology

Maria-Rosa Lloret

Universitat de Barcelona

25.1. Introduction

Oromo, formerly known as Galla, is a Cushitic language of the Afroasiatic family. Within Cushitic, Oromo, as well as Somali, Saho-Afar, Arbore-Dasenech, Burji-Deleba, Boni, and Rendille, belongs to the Lowland East Cushitic group. It is spoken in the southern part of Ethiopia (where together with Amharic it is the most important language) and northern Kenya, with native speakers living in Somalia and Sudan, too. Though it is one of the most widely spoken languages of Africa with upwards of seven million speakers (Heine et al. 1981), some figures mentioning more than ten (Gragg 1976) and up to twenty (Ali & Zaborski 1990) million speakers are considered high (Bender & Eteffa 1976).

25.2. Characteristics of the present study

The present study sets out a description of the phonological system of Oromo. It includes a presentation of the sounds and phonemes of the language and the way they tend to combine, as well as a description and classification of regular and contextually well-defined processes of sound change. Suprasegmental patterns are described in less detail, since this is a field which deserves further investigation. Issues strictly related to specific morphological processes, such as reduplication and lengthening/shortening of final segments concomitant with certain morphemes (i.e., length as a component of a specific morpheme), will not be described. The work benefits from previous descriptions of Oromo (see *References*, among others) and tries to systematically put together their findings, especially on issues about which scholars are most at variance.

On the subject of dialect differentiation, Oromo scholars agree on the mutual intelligibility between them, though, with the only exception of Stroomer's (1987) major dialectal study on the Southern area, the status and boundaries among the various dialects have not been clearly established.

Oromo studies usually refer to three major groups, namely Western, Eastern, and Southern. The Western group roughly includes Northern, Western, and Central dialects (cf. Heine 1981: 15); however, sometimes a large transitional area (cf. Gragg 1976: 172) or even a further Northern group (cf. Bryan 1948: 21–22) are distinguished. In this study, the description is based on the Western dialects, among which Mechaa (spoken in the Wellegga area) is the major representative. Dialectal variation will be pointed out whenever significant correspondences can be established through the descriptions of previous works.¹

Since the present study is descriptive, the reader will not find methodologically sophisticated analyses of the phonological processes. During the description, a distinction will be made between phonetics and phonemics. Phonemic notation will only include those sounds which are distinctive in the language, i.e., may cause differences in the meaning. The characterization of the phonological processes will be done on the grounds of the regularities observed from the phonetic data. Abstract representations will be disregarded. Nevertheless, whenever Oromo provides interesting data for linguistic research on the nature of phonological representations, it will be pointed out.

25.3. Segmental phonology

Oromo has the vowels and consonants shown in Tables 25-1 and 25-2.

Table 25-1. Oromo Vowels

i, ii	u, uu
e, ee	o, oo
a, aa	

In order to facilitate the exposition some simplified symbols are used. A list of the corresponding IPA phonetic symbols and the corresponding symbols in Latin script as used by most of the Oromo is presented in Table 25-3.

1. Throughout this work, the description of the Western dialect is based on the Mechaa variety as described in Gragg (1976, 1982) and Lloret (1988, 1989); the description of Eastern is based on the Harar variety as described in Owens (1980, 1985) and Ali & Zaborski (1990), and the description of Southern is based on the Orma, Waata, and Boraana varieties as described in Owens (1980), Heine (1981), and Stroomer (1987).

Table 25-2. Oromo Consonants

	Labial	Dento-Alveolar	Palatal	Velar	Laryngeal
Stops					
Voiceless		t	č	k	
Voiced	b	d	ǰ	g	
Ejective	p'	t'	č'	k'	'
Implosive		d'			
Fricatives					
Voiceless	f	s	š	(x) ^a	h
Sonorants					
Nasal	m	n	ñ		
Liquid		l r			
Glide	w		y		

^a/x/ occurs as a phoneme in Eastern Oromo, which corresponds to /k/ in other dialects. It might also occur as allophone of /k/ in other varieties (see the description of /k/ in § 25.3.2).

Table 25-3. Oromo Transcriptions

	IPA	Latin Script
c corresponds to	t̪	c
p'	p'	ph
č'	t̪'	ch
'	ʔ	'
š	ʃ	sh
ǰ	ɟ͡ʒ	j
t'	t'	th
k'	k'	kh
d'	ɗ	dh
ñ	ɲ	ny

Geminate ejectives and implosives are transcribed with glottalization in the second consonant only, e.g. /tʔ/ or /dʔ/. The same holds of heterorganic ejective clusters, e.g. /ptʔ/.

25.3.1. Vowels

Oromo makes a phonemic distinction between five short and five long vowels, which occur either within the word or at the end of the word. The minimal pairs in (1) exemplify this two-way distinction. Sequences of non-identical vowels do not occur.²

- | | | | | |
|-----|---------------|------------------|----------------|----------------|
| (1) | <i>d'aba</i> | 'I am unable' | <i>d'aaba</i> | 'I plant' |
| | <i>egaa</i> | 'then' | <i>eegaa</i> | 'watch! (pl.)' |
| | <i>d'isa</i> | 'I stretch skin' | <i>d'iisa</i> | 'I leave' |
| | <i>boruu</i> | 'east' | <i>booruu</i> | 'muddy' |
| | <i>d'ufii</i> | 'coming' | <i>d'uufii</i> | 'fart' |

Unstressed short vowels tend to be laxer and more centralized than their longer counterparts. This distinction is especially clear for the non-rounded series /a, e, i/, while the phonetic difference for the rounded vowels /o, u/ tends to be purely quantitative.³

- | | | | | | |
|-----|--------------|----------------|---------------|----------|-----------|
| (2) | <i>nama</i> | [námə] 'man' | <i>namni</i> | [námni] | 'man+NOM' |
| | <i>boruu</i> | [borúu] 'east' | <i>d'ufii</i> | [d'ufii] | 'coming' |

This short/long vowel contrast is phonetically realized both within a word and at the end of a word when it is not followed by a pause. In pre-pausal position there is also a two-way vowel contrast, but this has generally been described as a contrast between voiceless vowels vs. short vowels with final glottal stop. There is, however, a lot of variation in the realization of these final vowels among speakers and dialects, especially in relation to short vowels, which are better described as phonetic tendencies.

2. Although this is true for all dialects morpheme-internally, non-identical abutting vowels exceptionally arise through morphophonemic changes in some dialects. Western dialects never show sequences of non-identical vowels. Eastern dialects, though, show diphthongs originated by morphophonemic changes (e.g. *hak'+sis+a* → *haisisa* or *heesisa* 'I make wipe'). Southern dialects show bisyllabic vocalic sequences also originated across morpheme boundaries (e.g. *mač'aw+e* → *mač'ae* 'he was drunk'). (These issues are further discussed in §§ 25.3.4 and 25.3.5.)

3. Stroomer (1987: 27) reports that this lax realization also holds of /o/ for the Southern dialects.

In pre-pausal position, long vowels are phonetically shortened and usually closed by a glottal stop,⁴ except original long vowels that should be lengthened through grammatical marking (3a).⁵

- (3) a. *gogaa* [goga'], pre-pausally 'skin'
 [gogaa], elsewhere
 (Cf. *gogaa* [gogaa], always 'skin+GEN')
 b. *namaa* [nama'], pre-pausally 'man+GEN'
 [namaa], elsewhere
 (Cf. *nama* 'man')

The realization of pre-pausal short vowels shows a larger number of phonetic surface contrasts, though the following tendencies have been observed:

- (i) Short vowels tend to be pronounced as reduced, i.e. less audible, unless they bear a grammatical low tone (4a).⁶ (Reduced vowels are indicated by superscript in the phonetic transcription.)
- (ii) /i/ goes a step further in the reduction and may be completely deleted at the end of a word, unless it is preceded by two consonants or it carries a low tone. In this case, it generally reduces (4b).
- (iii) All short vowels may be pronounced with different degree of voicing according to the previous consonant: they tend to be voiced or partially devoiced when the preceding consonant is voiced but fully devoiced when the preceding consonant is voiceless (4c).⁷
- (iv) Reduced vowels tend to be pronounced as voiceless, while full short vowels are not (4d).

4. In Southern Oromo, these pre-pausal long vowels may be realized as short or as vowel plus laryngeal consonant plus (reduced) voiceless vowel, e.g. *garaa* [gara], [garaʔ], [garahʔ] 'belly' (cf. Stroomer 1987: 28).

5. On the possible interactions between the realization of final vowels and grammatical marking, see Stroomer (1987) for Southern Oromo and Lloret (1988, 1989) for Western Oromo. In the Orma and Waata Southern dialects, these grammatically lengthened long vowels may dissimilate, e.g., *hattuu* 'thief', *hattua* 'thief+GEN', originating one of the exceptional cases of non-identical vocalic sequences that occur in the language.

6. This does not always hold of the Harar Eastern variety, where pre-pausal short vowels are usually pronounced like long vowels, i.e. with a final glottal stop (cf. Owens 1985: 10, Ali & Zaborski 1990: xxii).

7. Though Harar tends to pronounce pre-pausal short vowels with a final glottal stop (see previous footnote), devoiced realizations also occur (cf. Owens 1985: 23).

- (v) Reduced vowels, especially if they are devoiced, tend to be dropped in normal connected speech (4e).⁸
- (4) a. *mana* [man^a] ‘house’, *mana* [manà] ‘it’s a house’
 b. *ilkaani* [ilkaan] ‘tooth’, *inni* [innⁱ] ‘he’, *ilkaani* [ilkaanⁱ] ‘it’s a tooth’
 c. *mana* [man^a], [man^a] ‘house’; *bofa* [bof^a] ‘snake’
 d. *mana* [man^a] ‘house’, *mana* [maná] ‘it’s a house’
 e. *man(a) guddaa sadii arge* ‘he saw three big houses’

In sum, the phonemic distinction that Oromo makes between short and long vowels is phonetically realized in a number of different ways, though in any case none of these allophones are to be considered phonemes, neither is there any reason to recognize a larger phonemic length contrast. The only change that systematically occurs is shortening of pre-pausal long vowels. The strongest tendency that has been observed is reduction of final /i/, which may or may not be completely dropped, depending on the dialect and the context in which it occurs.

25.3.2. Consonants

In general, consonants occur single in initial position while intervocally they may occur single, geminated, or as members of biconsonantal clusters. Phonemically, all Oromo words end in a vowel, though, as previously mentioned, this vowel may be phonetically dropped under certain conditions. Restrictions on the distribution of some consonants as well as their allophonic variation are presented below.

All consonants occur initially and medially, except /ʔ/ which never occurs initially (although it is true that a prothetic [ʔ] is added to vowel-initial words in absolute initial position) and /h/ which, except in Eastern Oromo, only occurs initially.⁹ /h/ also has the peculiarity that in some, but not all, words it can be dropped and, according to Oromo scholars, under no clear conditioning (5a) (cf. Gragg 1976: 174, Stroomer 1987: 16). The interchangeability of /h/ with zero seems to be a matter of idiolect (cf. Stroomer 1987: 16).

8. In this respect, there is also quite a big amount of dialectal variation, i.e. final unstressed non-low-toned /a/ tends to be deleted in the Arussi Eastern dialect but maintained in the Harar Eastern variety (cf. Owens 1985). It also usually drops in the Boraana and Waata Southern varieties (cf. Owens 1980, Heine 1981), and less frequently in the Western dialects.

9. The phonemic status of the glottal stop has been questioned, since in some dialects the two laryngeal consonants are in total complementary distribution and thus their occurrences are fully predictable (cf. Stroomer 1987: 52). This does not seem to be the case of Western and Eastern dialects, though it is true that its full phonemic status is at least problematic and deserves further research.

Thus, it is attributed to the lexicon, not to the phonemic inventory of the language. Only one minimal pair of initial /h/ versus initial vowel has been found (5b) (cf. Stroomer 1987: 16).

- | | | | |
|-----|----|---|--|
| (5) | a. | <i>harree</i> , * <i>arree</i> ‘donkey’ | <i>harka</i> , <i>arka</i> ‘hand’ |
| | | <i>hima</i> , * <i>ima</i> ‘I tell’ | <i>hirkoo</i> , <i>irkoo</i> ‘back’ |
| | b. | <i>hafa</i> ‘I remain behind’ | <i>afa</i> ‘I spread out
(to dry in the sun)’ |

All consonants except /h/ can occur single or geminate. In the Western and Southern dialects, this follows from the fact that geminate consonants only occur in medial position and /h/ only occurs initially. Geminate /ʀ/ only occurs in one instance and only in Western Oromo, i.e. (*h*)*oʀuu* ‘to be warm’, which corresponds to geminate /w/ in other dialects, i.e. *owwu*. In Southern Oromo, there are no instances of geminate /ʃ/, /ñ/, and /p/ either (cf. Stroomer 1987: 11). The following list of minimal pairs illustrates the opposition of short versus long consonants.

- | | | |
|-----|------------------------------------|--|
| (6) | <i>badaa</i> ‘red soil under fire’ | <i>baddaa</i> ‘highland’ |
| | <i>butaa</i> ‘snatcher’ | <i>buttaa</i> ‘period of time
of the Gada system’ |
| | <i>hod’a</i> ‘I suck’ | <i>hodd’a</i> ‘I sew’ |
| | <i>diree</i> ‘necklace’ | <i>dirree</i> ‘hill’ |
| | <i>kalee</i> ‘kidney’ | <i>kallee</i> ‘child’s garment’ |
| | <i>d’owa</i> ‘valley’ | <i>d’owwa</i> ‘I avoid’ |

In addition to the consonant phonemes presented at the beginning of § 25.3, some Oromo words contain /z/ and /p/, but they clearly are loans. In general, loanwords which have /z/ tend to be assimilated in the language as /s/ (occasionally also as /ʃ/) (7a), while /p/ tends to be assimilated as /f/ (occasionally also as /p’/ or /b/) (7b). More recently, some loans with /v/ have been introduced in the language, which also tend to be assimilated as /f/ (7c).

- | | | |
|-----|----|---|
| (7) | a. | <i>muuza</i> , <i>muusa</i> ‘banana’ (from Arabic, Amharic <i>muz</i>)
Cf. <i>aʃaʃa</i> ‘I command’ (Western) (from Amharic <i>azzázä</i>) |
| | b. | <i>poolisii</i> , <i>foolisii</i> ‘police’ (from Amharic <i>polis</i> , ultimately of European origin)
Cf. <i>p’eesaa</i> , <i>beesee</i> ‘money’ (Southern) (from Swahili <i>pesa</i>) |
| | c. | <i>vitaamini</i> , <i>fitaamini</i> ‘vitamin’ (from English) |

[p] also occurs as an allophone of /b/ when it is followed by a voiceless consonant (see § 25.3.7).

- (8) *obs+a* [opsa] (Western) 'I endure', *obsa* [opsa] (Southern: Boraana) 'patience'
č'ab+sa [č'apsa] (Western, Southern) 'I break' (transitive)
 Cf. *č'ab+a* [č'aba] (Western, Southern) 'I break' (intransitive)
č'ap'+a [č'ap'a] (Eastern, occasionally in Southern) 'I break' (intransitive)

/p/ never occurs in initial position, except in some loans which show /p/ in the original language. It may appear as /b/, too.

- (9) *p'eesaa* (Southern: Orma), *beesee* (Southern: Boraana, Waata) 'money' (from Swahili *pesa*) (cf. Stroomer 1987: 20)

Oromo shows the peculiarity of displaying a stop system which has /p/ and /b/, but a gap for /p/. According to universal typologies, this is a rare occurrence since glottalized segments are functionally the weaker members of the stop series and thus one would expect a system with a gap in the ejective series but a complete simple series (cf. Gamkrelidze 1978: 17). Oromo does follow universal claims in that among voiceless stops a gap in the labial position is natural because /p/ is the marked member of the series (cf. Gamkrelidze 1978: 15). Oromo also patterns with other Afroasiatic languages in having the absence of /p/ but the presence of the corresponding voiceless fricative /f/. Thus, one can look at this peculiarity of the labial series from a comparative point of view and claim that /f/ fills the labial gap, so that one ends with a three-way contrast *f/b/p'*. Under this interpretation, the absence of /p/ is no longer to be considered such a rarity.

/d/ is more or less retroflex in most dialects.¹⁰ In some lexical items, /d/ alternates with zero. This alternation is lexically restricted and may vary according to the dialect, e.g.

- (10) a. *d'uumma*, *uumma* (Western) 'peninsula' (cf. Gragg 1982: 132)
 b. *d'agamsa*, *agamsa* (Southern: Boraana) '*Carissa edulis*' (cf. Stroomer 1987: 16)
 c. *d'arga* (Southern: Orma, Waata), *arga* (Western and Southern: Boorana), *arka* (Eastern) 'I see' (cf. Stroomer 1987: 15, Gragg 1982: 19, Owens 1985: 254)
 d. *d'ameelaa* (Southern: Boraana), *ameeloo* (Southern: Orma) 'milk container' (cf. Stroomer 1987: 16)

10. In Andrzejewski's (1957: 355–56) study on the Boraana Southern dialect, /d/ is described as fully implosive without retroflexion. Stroomer (1987: 15), however, emphasizes its retroflex articulation in all the Southern varieties and even doubts of its glottalized character. He also points out that it may have a flap pronunciation between vowels.

The fact that Oromo has /t'/ as well as /d'/ is another rarity, because if a language displays an ejective series it usually does not have implosives at the same point of articulation. In fact, most other Cushitic languages only present either the implosive/retroflex segment (e.g. Saho-Afar, Somali, Konso) or the ejective one (e.g. Hadiyya). Only one other Cushitic language presents both segments, namely Dullay (Gawwada and Gollango). (Cf. Sasse 1979.)¹¹

/ʃ/ is not widely distributed. It alternates with [s] in a good number of words (11a).¹² In some lexical items, this alternation is dialectally determined (11b).

- (11) a. *iʃeeni*, *iseeni* (Western) 'she'
 biʃaani, *bisaani* (Western) 'water'
 b. *iʃiini* (Southern: Orma, Waata), *isiini* (Southern: Boraana) 'she'
 ʃeefitii (Southern: Boraana), *seeftii* (Southern: Orma) 'knife' (cf. Stroomer 1987:20)

/ʃ/ is found in initial and intervocalic position. It occurs initially in relatively few words. It does not occur in syllable-final position except in exceptional cases, i.e., in some loans and interjections where it always occurs geminated (12a) or as a result of lexical reduplication (12b).

- (12) a. *aʃʃabo* 'salt' (from Amharic)
 iʃʃoo 'bravo!'
 aʃʃaami 'hello!'
 b. *buʃbuʃii* 'very thick beer' (Western)

/ñ/ and /y/ are not very frequent either. They also occur initially in few words, especially the nasal. Neither of them, unless geminated, phonetically occurs in syllable final position. Moreover, there are relatively few cases of geminate /ñ/ or /y/. /č/ is not found in initial position except in one interjection (i.e., *čee* 'command to donkeys') and in some loans (e.g., *čiiḥa* 'chief', *čaačii* 'church', both from English). Unlike these palatals, /j/ and /č/ are widely distributed and occur in a large number of words. /čč/ is fairly common, though in nearly all cases it arises from morphophonological alternations (see § 25.3.4.8).

11. Within the Chadic family, implosives and ejectives at the same point of articulation are attested in Goemai and related languages.

12. The shift from /ʃ/ to /s/ is a general tendency of the total Ethiopian area (cf. Sasse 1975: 261–62). In some dialects, e.g. Boraana, /ʃ/ has been almost eradicated in native words. In the peripheral dialects it remains here and there vacillating with [s]. This has led to the replacement of /ʃ/ by /s/ in individual dialects during the last few centuries.

/k/ sometimes spirantizes into [x]. In Western Oromo, it is never spirantized. Within the Southern dialects, /k/ becomes [x] in initial and intervocalic position commonly in the Boraana variety but rarely in Orma and Waata (cf. Stroomer 1987: 17). In Eastern Oromo, however, /k/, with very few exceptions, occurs only geminate or as the second consonant of a cluster. In all other contexts [x] occurs, which has phonemic status in this dialect (cf. Owens 1985: 15).

- (13) *muka* ‘tree’: *mu[k]a*, always in Western, commonly in
Orma and Waata, occasionally in Boraana
mu[x]a, commonly in Boraana, occasionally in
Orma and Waata
muxa ‘tree’: *mu[x]a*, always in Eastern

25.3.3. Consonant clusters

Oromo has abutting consonants in medial position. All the sequences are bi-consonantal, because the language does not allow sequences of more than two consonants. The syllable break is almost always between the two consonants. In Western Oromo and in the Waata Southern variety there are few exceptions and they all occur in loanwords (14a). Harar Eastern and Boraana and Orma Southern varieties also show occurrences of monosyllabic /br/ clusters, which correspond to bisyllabic /rb/ and /rf/ clusters in Western and Waata Southern dialects, respectively; in Western Oromo monosyllabic [br] occasionally occurs in free variation with [rb] (14b). Other instances of /l/ occurring after labials /b, f/ are reported for Eastern and Southern dialects. These sequences are pronounced with a syllable break between the two consonants in careful speech but this break is fairly maintained in connected normal speech; Western dialects occasionally show /lb/ sequences in free variation with [bl] (14c).

- (14) a. *kristaana* ‘Christian’ (European origin)
kabriitii ‘match’ (from Amharic)
b. *dubra* (Eastern; Southern: Boraana, Orma; occasionally in
Western) ‘girl’
Cf. *durba* (usually in Western), *durfa* (Southern: Waata)
c. *kofla* (Southern), *xofla* (Eastern) ‘I laugh’
ablee (Eastern, Southern, occasionally in Western) ‘knife’
Cf. *kolfa* (Western), *albee* (usually in Western)

Most of the consonant clusters that appear within a morpheme in non-loanwords involve at least a liquid or a nasal in first position. Clusters where

the first consonant is /n/ are phonetically homorganic (see § 25.3.7.1).

- (15) *leenč'a* [leeñč'a] 'lion'
danfa [damfa] 'it boils'

/ñ/ shows a strong restriction on its distribution: it only exceptionally occurs in some Western items, where it exclusively combines with a following glottal stop and alternates in free variation with a geminate palatal sequence (e.g., *mañ'ee*, *maññee* 'lower part of leg'). The restricted distribution of /ñ/ follows from a strong restriction that holds of all palatal consonants, namely a palatal consonant cannot be followed by a consonant: /š, č, č'/ never occur as the first element of a non-geminate consonantal cluster;¹³ /y/, like /ñ/, exceptionally occurs in few Western items combining with a glottal stop and alternating with geminate sequences, e.g. *bay'ee*, *bayyee* 'much, many'); /y/ also exceptionally occurs in the Waata Southern variety, e.g. *toyba* 'seven' (cf. *torba* in Western, Eastern, and Boraana Southern, and *tolba* or *tolp'a* in Orma Southern) (cf. Stroomer 1987: 25).

Among other possible consonant clusters, it is important to note the frequent occurrence of the sequence /f/ plus consonant. This has a straightforward explanation in Oromo, since /s/ regularly changed into /f/ when immediately preceded by another consonant. This labialization morphophonemically still applies in most dialects (see § 25.3.4.1). The sequence /s/ plus consonant is only found in loanwords, e.g.

- (16) *maskootii* 'window' (from Amharic)
masgiida (Western), *meskitii* (Southern) 'mosque' (from Arabic)
misraaččoo (Western) 'good news' (from Amharic)

Other possible combinations do not occur often and very few examples are found, some of which just occur in one of the varieties: /kt/, /ks/, /k's/, /kn/, /bs/, /bd/, /bj/, /gd/, and /gn/ in Western Oromo (cf. Lloret 1988: 22–23); /bs/ and /bj/ in Boraana, /dy/ in Waata, and /n/ plus /y/ in all Southern dialects (cf. Stroomer 1987: 25).

25.3.4. Morphophonemic changes involving consonant clusters

Consonant clusters across morpheme boundaries arise only between the final consonant of a stem and /n/, /t/, or /s/, or between /n/ and a consonant-initial stem. The first two cases originate within both verbal and nominal morphology while the last two only originate within verbal morphology.

13. /š/ exceptionally occurs in preconsonantal position in one lexically reduplicated word, i.e., *bušbušii* 'very thick beer' (Western).

Some changes are grammatically restricted since they only occur in isolated morphemes; others are unrestricted, i.e., the morphophonological changes involved are independent of the grammatical status of the forms. Changes involving stems ending in a glide or in a laryngeal consonant will be independently treated (see § 25.3.5), because they are quite complex and show a large amount of dialectal variation.

Oromo has two changes that specifically occur in nominative forms. In Oromo, nouns and descriptive adjectives have distinct nominative forms, which are formed by suffixation of a marker /ni/ (/ti/ in some feminine forms ending in a short vowel).¹⁴ When this suffix is attached to a form ending in a long vowel, no change occurs (17a); however, nouns and adjectives ending in a short /a/ drop this vowel (17b).

- (17) a. *saree+ni* → *sareen(i)*¹⁵ 'dog+NOM'
 adii+ni → *adiin(i)* 'white+NOM'
 b. *nama+ni* → *namni* 'man+NOM'
 lafa+ti → *lafti* 'earth+NOM'

The nasal of the nominative suffix may optionally assimilate to the point of articulation of a preceding /m/.

- (18) *nam(a)+ni* → *namni*, *nammi* 'man+NOM'
 č'oom(a)+ni → *č'oomni*, *č'oommi* 'fat+NOM'
 Cf. *tum+na* → *tumna*, **tumma* 'we beat'

Unlike these two changes, which are restricted to the nominative suffix,¹⁶ other morphophonological changes are unrestricted, though some are only found within verbal morphology because, as previously mentioned, some environments do not originate within nominal morphology. Only changes which can be ascertained not to occur in nominal morphology will be considered to be restricted to verbal forms.

14. In order to facilitate the exposition, /ni/ and /ti/ are considered the basic nominative suffixes, though variation in vowel length and combinations of both (e.g. *ti(i)ni*) are reported for specific varieties (cf. Owens 1985: 100–2, Stroomer 1987: 167, Lloret 1988: 116–18).

15. On the phonetic realization of this final /i/, which bears a high tone, see § 25.3.1. In the examples, this vowel is given between parentheses when it usually drops.

16. A further morphophonemic restricted change could be considered to occur in nominative forms, namely /n/ drops after an abutting consonantal cluster, e.g. *harka+ni* → *hark+ni* → *harki* 'hand+NOM', *duumessa+ni* → *duumess+ni* → *duumessi* 'cloud+NOM' (cf. Gragg 1976: 183, Lloret 1988: 118). This analysis, though, presupposes a certain degree of abstraction, since these cases can also be interpreted as forms that merely attach a separate nominative suffix /i/ (cf. Owens 1985: 101, Stroomer 1987: 167, Banti 1988: 47).

25.3.4.1. *s-Labialization*

/s/ becomes [f] when it is followed by a consonant. This change always takes place in Western and Eastern Oromo and in the Boraana and Orma Southern varieties, though the alternation with [s] is optional (19a). In the Waata Southern dialect, however, /s/ tends to be either maintained or (partially) assimilated to the following consonant (19b).

- (19) a. **Western, Eastern, and Boraana and Orma Southern varieties:**
aḵḵees+na → *aḵḵeefna* (occasionally: *aḵḵeesna*) ‘we kill’
čḵiis+ta → *čḵiifta* (occasionally: *čḵiista*) ‘you lie down’
- b. **Waata Southern variety:**
huk'+s+ta → *huk'ista*, *huk'ibta*, *huk'itta* ‘you make thin’
gudd+s+na → *guddisna*, *guddimna* ‘we raise children’

25.3.4.2. *Nasal Total Assimilation*

/n/ completely assimilates to a preceding liquid (20a) and to a following liquid or glide (20b).¹⁷ In Southern Oromo, /n/ also assimilates to a following /s/ (20c).

- (20) a. *gaal(a)+ni* → *gaalli* ‘camel+NOM’
gal+na → *galla* ‘we enter’
moor(a)+ni → *moorri* ‘fat+NOM’
mur+na → *murra* ‘we cut’
- b. *hin+latu* → *hillatu* ‘it does not sprout’
hin+rafu → *hirrafu* ‘he does not sleep’
hin+wareegu → *hiwwareegu* ‘he does not promise’
hin+yaadu → *hiyyaadu* ‘he does not think’
- c. *hin+sirba* → *hissirba* ‘he does not dance’ (Southern dialects only)

25.3.4.3. *Glottalization*

/t/ assimilates in ejection to a preceding consonant. In the case of a preceding velar, other changes may occur (see § 25.3.4.6) (21a). In the Southern varieties, the ejection is occasionally lost (21b).

- (21) a. *č'ap'+ta* → *č'apt'a* (Eastern, occasionally in Southern) ‘it breaks’
 Cf. *č'ab+ta* → *č'abda* (Western, usually in Southern)
fit'+ta → *fitt'a* (Eastern, Western)¹⁸ ‘you enter’
hiik'+ta → *hiikt'a* (Western) ‘you go’; *hiitt'a* (Eastern)

17. It should be pointed out that Stroomer (1987: 41) does not mention assimilation of *hn* to a following glide for the Boraana Southern dialect.

18. In the Southern varieties *h'* is treated as a double consonant (see § 25.3.4.9).

- b. *d'ek'+ta* → *d'ett'a*, *d'etta* (Southern: Boraana, Orma) 'you go'; *d'eča* (Southern: Waata)

In Western Oromo, ejection also spreads over the second consonant in the case of an original /t'+s/ sequence, which becomes [čč'] through a palato-affrication process described below.

- (22) *lit'+siis+a* → *fičč'iisa* (Western) 'I make enter'
Cf. *t'ut'+siis+a* → *t'usiisa* (Eastern) 'I make suck'

25.3.4.4. Voice assimilation between stops

/t/ assimilates in voicing to a preceding stop. In the case of a preceding velar, other changes may occur (see § 25.3.4.6) (23a). This change does not affect /d'/, which behaves like a voiceless consonant and does not induce voice assimilation (see § 25.3.4.6). In the Southern varieties, the voiceless character of /t/ is sometimes optionally maintained: it is never maintained when it is preceded by a coronal, but it may be maintained in all dialects after a labial stop, [bt], and in Orma and Waata after a velar stop (23b) (cf. Stroomer 1987: 46).

- (23) a. *did+ta* → *didda* 'you refuse'
k'ab+ta → *k'abda* 'you have'
k'adda (occasionally in Waata Southern variety)
jab+tuu → *jabduu* 'strong (fem.)'
jadduu (occasionally in Waata Southern variety)
d'iig+ta → *d'iigda* (Western) 'you bleed'
d'iidda (Southern: Boraana, Orma)
d'iyyda (Eastern)
d'iija (Southern: Waata)
mid'aag+tuu → *mid'aadduu* (Southern: Boraana, Orma) 'nice (fem.)'
mid'aaɣuu (Southern: Waata)
jaɣ+ta → *jaɣɣa* (Western) 'you boast' (no examples reported in other dialects)
- b. *did+ta* → **didta*
k'ab+ta → *k'abta* (occasionally in Southern)
d'iig+ta → *d'iitta* (occasionally in Orma Southern variety)
d'iica (occasionally in Waata Southern variety)¹⁹

19. Stroomer (1987: 44) points out that /g+t/ becoming [č] instead of [j] might be explained in terms of analogy to other verbs. In general, no example with voiceless alternant is reported within nominal morphology (e.g. *deg(a)+ti* → *deɣi*, **deči* 'people+NOM', in Waata).

25.3.4.5. *Palatal assimilation*

A consonant preceded by a stem ending in a palatal consonant becomes palatal, too (24a). It may undergo further changes, i.e., a coronal totally assimilates to a following nasal (24b) while /t/ completely assimilates to a preceding palatal (24c). Very few examples are found because of the limited distribution of palatals.

- (24) a. *d'eereeñ(a)+ni* → *d'eereeñni* (Eastern) 'tallness+NOM'
 b. *jaǰ+na* → *jañña* (Western) 'we boast'
 c. *jaǰ+ta* → *jaǰja* (Western) 'you boast'

In the Waata Southern variety, a velar stop plus /n/ becomes [ɲ]. This process is probably related to the palatalization induced by a velar stop over a preceding vowel in Eastern Oromo. (See § 25.3.4.6.)

- (25) *beek+na* → *beeña* (Southern: Waata) 'we know'
d'ek'+na → *d'eña* (Southern: Waata) 'we go'
neek'(a)+ni → *neeñi* (Southern: Waata) 'lion+NOM'
d'ug+na → *d'uña* (Southern: Waata) 'we drink'
lag(a)+ni → *lañi* (Southern: Waata) 'river+NOM'

25.3.4.6. *Stop consonant assimilation*

There is a large amount of variation among dialects with respect to the degree of assimilation allowed between a stop and a following /t/ or /n/. The only change that systematically occurs in all dialects is total assimilation of a non-glottalized coronal stop to a following nasal, though examples with final /j/ are only reported for Western Oromo (26). If a stop is followed by /t/, voice assimilation applies as previously described, cf. (23).

- (26) *fid+na* → *finna* 'we bring', *haad(a)+ni* → *haanni* 'rope+NOM'
bit+na → *binna* 'we buy', *hantuut(a)+ni* → *hantuunni* 'rat+NOM'
jaǰ+na → *jañña* (Western) 'we boast'

/d/ completely assimilates to /t/ and /n/ in Eastern and Southern Oromo; in the latter, the previous vowel may be short, in which case it lengthens (27a). /t/ only assimilates in the Eastern varieties (27b).

- (27) a. *fuud'+ta* → *fuutta* 'you take', *haad'+ti* → *haatti* 'mother+NOM'
 (Eastern, Southern)
fed'+na → *feenna* (Southern) 'we want'
feed'+na → *feenna* (Eastern)
 Cf. *fuud'+ta* → *fuuta*, *haad'+ti* → *haati*, *fed'+na* → *feena* (Western)

- b. *fiit'+na* → *fiinna* (Eastern) 'we finish'
 Cf. *fiit'+na* → *find'a* (Western), *fiit'ina* (Southern)

If a velar stop precedes /t/ or /n/ no change other than voice assimilation and glottalization occurs in Western Oromo. However, in Eastern and in Boraana and Orma Southern, the stop completely assimilates to the following consonant. (Voice assimilation and glottalization take place as previously described.) In the case of the Eastern dialects, assimilation of the velar consonant induces a strong palatalization of the preceding vowel, except in the case of /ee/ (cf. Owens 1985: 24). In the Orma Southern variety, /k+t/ may voice into [dd] while [gd] is occasionally maintained (cf. Stroomer 1987: 41). In Boraana, /k'+n/ exceptionally changes into [nd']²⁰ (cf. Stroomer 1987: 39).

- (28) *beex+ta* → *beetta* (Eastern) 'you know'
beex+na → *beenna* (Eastern) 'we know'
mux+ni → *muynni* (but also: *muxni*) (Eastern) 'tree+NOM'
beek+ta → *beetta* (Southern: Boraana, Orma), *beedda* (Orma) 'you know'
beek+na → *beenna* (Southern: Boraana, Orma) 'we know'
muk(a)+ni → *munni* (Southern: Boraana, Orma) 'tree+NOM'
d'uug+ta → *d'uuydda* (Eastern) 'you drink'
d'uug+na → *d'uuyinna* (Eastern) 'we drink'
lag(a)+ni → *laynni*, *leenni* (Eastern) 'river+NOM'
d'ug+ta → *d'udda* (Southern: Boraana, Orma); *d'utta*, *d'ugda* (Orma) 'you drink'
d'ug+na → *d'unna* (Southern: Boraana, Orma) 'we drink'
lag(a)+ni → *lanni* (Southern: Boraana, Orma) 'river+NOM'
d'ak'+ta → *d'aytt'a*, *d'eett'a* (Eastern) 'you arrive'
d'ak'+na → *d'ayinna*, *d'eenna* (Eastern) 'we arrive'
warak'(a)+ni → *warayinna*, *wareenna* (Eastern) 'paper+NOM'
d'ak'+ta → *d'att'a*, *d'atta* (Southern: Boraana) 'you go'
d'ek'+ta → *d'ett'a*, *d'etta* (Southern: Orma) 'you go'
d'ak'+na → *d'anna*, *d'and'a* (Southern: Boraana) 'we go'
d'ek'+na → *d'enna* (Southern: Orma) 'we go'
waak'(a)+ni → *waanni* (Southern: Boraana, Orma) 'god+NOM'

In Southern Oromo, /b/ and /f/ may optionally assimilate, either completely or partially, to a following /n/.

20. This change to [nd'] is only found in verbal forms in Stroomer (1987).

- (29) *k'ab+na* → *k'abna* (Orma) 'we have'
 k'amna (Boraana, Orma, Waata)
 k'anna (occasionally in Waata)
 arrab(a)+ni → *arrabni* (Orma) 'tongue+NOM'
 arramni (Boraana, Orma, Waata)
 arranni (occasionally in Waata)
 d'uf+na → *d'ufna* (Boraana, Orma) 'we go'
 d'umna (Boraana, Waata)
 d'unna (occasionally in Waata)
 bof(a)+ni → *bofni* (Boraana, Orma) 'snake+NOM'
 bomni (Boraana, Waata)
 bonni (occasionally in Waata)

In the case where a stop precedes /s/, which can be only (part of) the causative verb extension, total assimilation only occurs with a velar stop and only in the Southern varieties, though in Waata /k'+s/ usually reduces to [s] (see § 25.3.4.7).

- (30) *beek+s+a* → *beessa* (Southern) 'I make know'
 d'ug+siis+a → *d'us(s)iisa* (Southern) 'I make drink'
 fiiik'+sa → *fiissa* (Boraana, Orma) 'I whistle'
 fiinsa (rarely in Boraana)
 fiisa (Waata)

25.3.4.7. Stop deletion

In Western Oromo, /d'/ is systematically dropped before a consonant. If the preceding vowel is short, it obligatorily lengthens (31a). This deletion occurs in all dialects when the following consonant is /s/ (31b). In Eastern Oromo, /t'/ also drops before /s/ (31b).

- (31) a. *fuud'+ta* → *fuuta* (Western) 'you take'
 haad'(a)+ti → *haati* (Western) 'mother+NOM'
 fed'+na → *feena* (Western) 'we wish'
 b. *fuud'+sis+a* → *fuusisa* 'I cause to take'
 c. *t'ut'+siis+a* → *tusiisa* (Eastern) 'I make suck'

In Eastern Oromo, a velar stop, which induces palatalization of the previous vowel, is dropped before /s/; in the Waata Southern variety /k'+s/ also reduces to [s], though some exceptions are found (see § 25.3.4.9) (32a). In all Southern varieties, /g+s/ may optionally reduce to [s] (32b).

- (32) a. *beex+sis+a* → *beesisa* (Eastern) ‘I make know’
d’uug+sis+a → *d’uuysisa* (Eastern) ‘I make drink’
hook’+s+a → *hooyisa* (Eastern) ‘I make scratch’
hoosa (Southern: Waata)
 b. *d’ug+siis+a* → *d’ussiisa* ‘I make drink’
d’usiisa (Southern)

25.3.4.8. Palato-affrication

In all dialects, a non-glottalized coronal plus /s/, which can only be (part of) the causative verb extension, becomes [čč] (Western, Eastern) or [č] (Southern). In Western Oromo, this also holds of /t/ (33a). In all dialects, /s/ becomes [č] after /l/ (33b).

- (33) a. *k’ot+sisiiis+a* → *k’oččisiisa* (Western, Eastern) ‘I make dig’
k’očisiisa (Southern)
did+sisiiiis+a → *diččisiisa* (Western, Eastern) ‘I make refuse’
dičisiisa (Southern)
aɣaɣ+siis+a → *aɣaččisiisa* (Western) ‘I make command’
lit’+siis+a → *ličč’iisa* (Western) ‘I make enter’
 b. *gal+s+a* → *galča* ‘I bring home’
bul+s+a → *bulča* ‘I spend the night’

In the Waata Southern dialect, a velar plus /t/ becomes affricate too. (The palatalization is probably related to the palatalization induced by a velar stop over the preceding vowel in Eastern Oromo, see § 25.3.4.6.)

- (34) *beek+ta* → *beeča* (Southern: Waata) ‘you know’
d’ek’+ta → *d’eča* (Southern: Waata) ‘you go to’
d’ug+ta → *d’uɣa, d’uča*²¹ (Southern: Waata) ‘you drink’
deg(a)+ti → *deɣi* (Southern: Waata) ‘people+NOM’

25.3.4.9. Vowel epenthesis

Oromo does not allow sequences of more than two non-syllabic segments. When such a sequence would arise, [i] is inserted between the second and third consonants.

- (35) *elmta* → *elmita* ‘you milk’
fayyna → *fayyina* ‘we are well’
gudd+s+a → *guddisa* ‘I raise (children)’
add(a)+ti → *additi* (Southern: Boraana) ‘forehead+NOM’ (*addi*, in other dialects)

21. See footnote 19.

In Southern Oromo, this also holds of stems ending in /č/, /č'/, /t'/, and optionally /d'/, even if the affricate is derived, i.e., [č] originated through palato-affrication (36a) (cf. Stroomer 1987: 54). In all dialects, there are exceptional forms in which [i] appears between the stem and the causative morpheme (36b).

- (36) a. *boč'+ta* → *boč'ita* (Southern) 'you carve wood'
t'uut'+na → *t'uut'ina* (Southern) 'we suck'
hod'+s+a → *hod'isa* (Southern) 'I make suck'
duud+s+ta → *duučita* (Southern) 'you fill up a hole'
- b. *damač'+siisa* → *damač'isiisa* (Western) 'I make crush'
huk'+s+ta → *huk'ista, huk'ibta, huk'itta* (Southern: Waata) 'you make thin'

In Eastern Oromo, [a] is inserted between the first and second consonants if the second consonant is a liquid (37a) (cf. Owens 1985: 21). This also occasionally occurs in Southern Oromo, where the epenthetic vowel may further copy the preceding vowel (37b) (cf. Stroomer 1987: 55).

- (37) a. *k'ofl+ta* → *k'ofalta* (Eastern) 'you laugh'
dabr+ta → *dabarta* (Eastern) 'you pass'
- b. *kofl+ta* → *kofalta* (Southern: Boraana, Waata)
kofl+ta → *kofolta* (Southern: Orma)

25.3.4.10. *Special changes*

A group of verbs, which are very frequent because of the middle voice extension, show unusual conjugated forms. The changes they involve are thus restricted to specific verbal morphemes. In prevocalic position, they show a geminate [dd'] in the first person singular and in the imperative (38a) and a geminate [čč] in the infinitive form²² (38b), while [t] occurs in the remaining prevocalic forms (38c). In preconsonantal position, this segment alternates like /t/ (38d).²³

- (38) a. *bit+a(a)dd'+a* 'I buy for myself'
bit+a(a)dd'+uu 'buy for yourselves!'
- b. *bit+a(a)čč+uu* 'to buy for one self'

22. Other [cc(')] forms have been reported as alternative infinitive forms for some verbs ending in alveolar obstruent, e.g. *fi'tuu, fičč'uu* 'to finish', *gamma'duu, gammačč'uu* 'to want', *god'uu, gočč'uu* 'to make' (cf. Moreno 1939: 73).

23. Oromo scholars have provided different analyses of these verbs, assuming either an original /dd'/ form (cf. Gragg 1976: 186, Stroomer 1987: 153) or an original /t/ form (cf. Owens 1985: 64). For a more abstract analysis see Lloret (1988: 187–99).

- c. *bit+a(a)t+a* ‘he buys for himself’
bit+a(a)t+an(i) ‘they buy for themselves’
- d. *bit+a(a)t+ta* ‘you buy for yourself’
bit+a(a)n+na ‘we buy for ourselves’ (cf. *bit+na* → *binna* ‘we buy’)
bit+a(a)č+čisiis+a ‘I make buy for myself’ (cf. *bit+sisiis+a* → *biččisiisa* ‘I make buy’)

There is only one attested case where the long implosive behaves like a geminate, namely the verb *hodd-* ‘sew’ (e.g. *hodd’ita* ‘you sew’, *hodd’a* ‘he sews’, *hodd’uu* ‘to sew’).

25.3.5. Morphophonemic changes involving glides and laryngeal segments

Stems ending in /w, y, ’, h/²⁴ undergo changes that involve these segments as well as the preceding vowels. In Western and Southern Oromo, a glide becomes [ʔ] if it is followed by a vowel-initial suffix. In the Southern varieties, the glide may also completely delete, which may originate heterorganic vocalic sequences.

- (39) *d’aw+a* → *d’a’a* (Western, Southern), *d’aa* (Southern) ‘I weave’
 Cf. *d’ow+a* → *d’owa* (Eastern) ‘I hit’
d’agay+e → *d’aga’e* (Western, Southern), *d’agae* (Southern) ‘I heard’
 Cf. *d’agay+e* → *d’ageye* (Eastern)

In Western and Southern Oromo, stems ending in /ʔ, w, d/ lengthen the previous vowel, if it is short, when they are followed by a consonant-initial suffix. The glottal stop further drops; the glide and the implosive may undergo and entail other changes. (See § 25.3.4.6 and § 25.3.4.7 for the behavior of /d/.)

- (40) *ta’+ta* → *taata* (Western, Southern) ‘you become’
d’aw+sis+a → *d’oofsisa* (Western), *d’ees(s)isa* (Southern) ‘I make weave’
fed’+na → *feena* (Western), *feenna* (Southern) ‘we want’

In Western Oromo, lengthening of the preceding vowel in pre-consonantal position also occurs if the stem ends in /y/ (41a). In Eastern Oromo, it does occur when the stem ends in a laryngeal consonant, which further deletes, and when /y/ is followed by /t/ or /n/, which undergo other changes (41b).

24. In medial position, /h/ only occurs in Eastern Oromo, which appears as // or zero in other dialects, e.g., *taha* (Eastern), *ta’a* (Western), *taa* (Southern) ‘I become’. Stems ending in /V/ are considered to end in /VV/ in Stroemer (1987: 51–2).

- (41) a. *d'agay+ta* → *d'ageessa* (Western) 'you hear'
d'agay+na → *d'ageeñña* (Western) 'we hear'
d'agay+sis+a → *d'ageessisa* (Western) 'I make hear'
- b. *tah+ta* → *taata* (Eastern) 'you become'
deebi'+sis+a → *deebiisisa* (Eastern) 'I make return'
d'agay+ta → *d'ageessa* (Eastern) 'you hear'
d'agay+na → *d'ageeñña* (Eastern) 'we hear'
dallay(a)+ni → *dalleeni* (Eastern) 'fence+NOM'

In Eastern Oromo, short /a/ may be affected by vowel harmony. It becomes [e] if the stem ends in a laryngeal consonant or in /y/ and the suffix begins in /e/ (42a). It becomes [o] if the stem ends in a laryngeal consonant and the suffix begins in /u(u)/; if the suffix has short /u/, it may further become [o] (42b). (Cf. Owens 1985: 21.)

- (42) a. *tah+e* → *tehe* (Eastern) 'he became'
ja'+e → *je'e* (Eastern) 'he said'
d'agay+e → *d'ageye* (Eastern) 'he heard'
- b. *bah+uu* → *bohuu* (Eastern) 'leaving (noun)'
ja'+u → *jo'u* (Eastern) 'let him say'
tah+u → *tohu, toho* (Eastern) 'he becomes (in subordinate clause)'

In preconsonantal position, /w/ systematically becomes [f] in Western Oromo (43a). In Eastern, preconsonantal /w/ assimilates to a preceding /o/; it optionally becomes [y] if the preceding vowel is /a/ and the next consonant is /t, n/; it deletes when it is followed by /s/ (43b). In the Boraana Southern variety, /w/ becomes [f] before /t/, [m] before /n/, and [s] before /s/, which may be optionally dropped; in the Orma Southern variety, /w/ assimilates in different degrees to the following consonant, which may be optionally dropped after /s/; in the Waata Southern variety, /w+t/ becomes [č], /w+n/ becomes [ñ], and /w+t/ becomes [s] (43c). In Southern and Eastern Oromo, /w/ after /s/ raises and fronts a preceding /a(a)/ into [e(e)], see (43b) and (43c).

- (43) a. **Western:**
d'aw+ta/na/sisa → *d'oofa/d'oofna/d'oofsisa* 'you/we/I make weave'
- b. **Eastern:**
d'ow+ta/na/sisa → *d'oota/d'oona/d'oosisa* 'you/we/I make hit'
beelaw+ta/na → *beeloyta/beeloyna* 'you/we become hungry'
 (optional)
beelaw+sa → *beelesa* 'I make hungry'

c. **Southern:**

d'aw+ta/na/sisa → *d'oofal/d'oomna/d'ees(s)isa* (Boraana) 'you/
we/I make hit'
d'oot(t)a/d'oon(n)a/d'ees(s)isa (Orma)
d'ooča/d'ooña/d'eesisa (Waata)

In preconsonantal position, /y/ always palatalizes a preceding /a(a)/ vowel, which becomes [e(e)]. In Western Oromo, /y+t/ further becomes [ss] and /y+n/ becomes [ññ] (44a). In Eastern, the same changes occur, which also affect final /aa'/ stems,²⁵ though in the Harar variety [ññ] only appears with- in verbal morphology (cf. Owens 1985: 20, Banti 1988: 47). The sequence /y+s/ becomes [ss] in Western while it simplifies into [s] in Eastern and South- ern (44a–c). In the Boraana and Orma Southern varieties, /y/ always deletes before a consonant; in Waata, /y+t/ becomes [č] and /y+n/ becomes [ñ] (44c).

(44) a. **Western:**

d'agay+ta/na → *d'ageessa/d'ageeñña* 'you/we hear'
d'agay+sisa → *d'ageessisa* 'I make hear'
nagay(a)+ni → *nageeññi* 'peace+NOM'

b. **Eastern:**

d'agay+ta/na → *d'ageessa/d'ageeñña* 'you/we hear'
taa'+ta/na → *teessa/teeñña* 'you/we sit'
d'agay+sisa → *d'ageesisa* 'I make hear'
nagay(a)+ni → *nageeññi* (Arussi) 'peace+NOM'
 Cf. *nagay(a)+ni* → *nageeni* (Harar) 'peace+NOM'

c. **Southern:**

d'agay+ta/na → *d'ageeta/d'ageena* (Boraana, Orma) 'you/we hear'
d'agay+ta/na → *d'ageeča/d'ageeña* (Waata)
d'agay+sis+a → *d'ageesisa* (Boraana, Orma, Waata) 'I make hear'
nagay(a)+ni → *nageeni* (Boraana, Orma), *nageeñi* (Waata)
 'peace+NOM'

25.3.6. *Vowel length alternation*

In Oromo, some inflectional and derivational suffixes exhibit alternation in vowel length depending on the length of the preceding vowel, namely the suffix has a long vowel if the vowel of the preceding syllable is short, and a short vowel if the vowel of the preceding syllable is long. This morpho-

25. These final /aa'/ Eastern stems significantly correspond to final /aay/ stems in other dia- lects, e.g., *taa'+ta* → *teessa* (Eastern), *taay+ta* → *teessa* (Western), *teeta* (Southern: Boraana, Orma), *teeča* (Southern: Waata) 'you sit'.

phonemic vowel alternation is mainly differentiative of Western Oromo (cf. Gragg 1976: 177, Lloret 1988: 77–81). In Eastern and Southern Oromo, it does not systematically apply though some cases are found, especially within the causative forms (cf. Owens 1985: 63, Stroomer 1987: 424).

- | | | | | |
|------|---------------------|---------------|------------------------|---------------------------|
| (45) | <i>ul+oota</i> | ‘sticks’ | <i>gaal+ota</i> | ‘camels’ (Western) |
| | <i>lugn+ooma</i> | ‘I am coward’ | <i>afčaal+oma</i> | ‘I am kind’ (Western) |
| | <i>tee(s)+sis+a</i> | ‘I make sit’ | <i>k’oč(č)i+siis+a</i> | ‘I make dig’ (all dials.) |

25.3.7. Unrestricted consonantal changes

Some changes that Oromo consonants undergo do not exclusively apply accross morphemes but whenever the phonetic context is satisfied.

25.3.7.1. Nasal Place Assimilation

/n/ assimilates to the point of articulation of a following consonant.

- (46) *leenč’a* [leeñč’a] ‘lion’
danfa [damfa] ‘it boils’
hinbaru [himbaru] ‘he does not learn’
hinkennu [hiŋkennu] ‘he does not give’

25.3.7.2. Glottal Stop Assimilation

A glottal stop may optionally assimilate to a preceding sonorant consonant or to a following vowel, e.g.

- (47) *bal’aa* [bal’aa], [ballaa] ‘wide’
bay’ee [bay’ee], [bayyee] ‘much’
(ka’+uu →) ka’uu [ka’uu], [kawuu] ‘to rise’
(d’agay+uu →) d’aga’uu [d’aga’uu], [d’agawuu] ‘to hear’

25.3.7.3. Stop Devoicing

A stop tends to devoice when it is followed by a voiceless consonant (48a). It also devoices before pause, when the final vowel is phonetically realized as voiceless (48b) (cf. Owens 1985: 23).

- (48) a. *obs+a* [opsa] (Western) ‘I endure’
obsa [opsa] (Southern: Boraana) ‘patience’
obsaa [opsaa] (Southern: Boraana) ‘officer’ (from English)
č’ab+sa [č’apsa] (Western, Southern) ‘I break’
k’ab+ta [k’apta] (occasionally in Southern) ‘you have’
 Cf. *k’ab+ta → k’abda, k’adda* (usually in Southern)

- fiig+sis*a [fiiksisa] (all dialects) 'I make run'
 b. *k'ab+ta* → *k'abda* [k'apt̪a] (in pre-pausal position) 'you have'
gub+ta → *gubda* [gupt̪a] (in pre-pausal position) 'you burn'

25.4. Suprasegmentals

Phonetically, Oromo contrasts high and low tones as well as different stress levels. Pitch and stress are assigned according to the phonological shape of a word in isolation and as a result of the grammatical information it encodes. Although there is general agreement on the tonal character of the language, as for almost all Cushitic languages (cf. Sasse 1981: 205), the exact status of tone is far from being clearly established. The large amount of dialectal variation that the language presents on this topic and the fact that tone on a word shows considerable variation according to the context in which the word occurs are the main problems for the description of the prosodic system of Oromo.

It is generally accepted that in Oromo tone is determined primarily morphosyntactically, and that its prosodic system conforms to pitch-accent systems, in the sense that tone is specified on a single syllable (usually on the penultimate) and from it the overall tonal pattern is identified on the morpheme. However, since in Oromo each morpheme contributes to the tonal pattern of the word, it differs significantly from other paradigmatic cases of pitch-accent languages, such as Japanese (cf. Owens 1985: 35–36). One further point to be mentioned is the correlation between tone and stress. In general, the stressed syllable is perceived as the first syllable of a word with high pitch (cf. Owens 1985: 37). Their complex relationship, though, seems to be more adequately established within the autosegmental framework, as Banti (1988) has done for the Arussi Eastern variety.²⁶ The following assumptions are fairly well accepted.

Verbal stems are toneless. They acquire tone by tonal patterns of the verbal affixes or by other tonal patterns of specific grammatical information. For instance, the affirmative imperative forms have a high–low tonal pattern in the singular, high being associated with the penultimate syllable, while a low–high pattern in the negative, high being associated with the penultimate syllable.

26. Banti (1988: 43) points out that the Oromo prosodic system involves a somewhat elaborate set of stress rules and a very simple set of tone rules. This statement presumably holds for other Cushitic languages, too.

- (49) *rafi* [ráfi] ‘sleep!’ (*hin+rafini* → *hirràfini* [hìrráfín(í)] ‘don’t sleep!’
taa’i [táa’i] ‘sit!’ *hintaa’ini* [hìntàa’ín(í)] ‘don’t sit!’

Most nouns (and adjectives) in isolation have a predictable tonal pattern according to their segmental shape, which may vary among dialects. Some of them, however, seem to be lexically marked. For instance, in Western and Eastern Oromo nouns ending in a long vowel have a low–high tonal pattern, low being associated with the penultimate syllable, though there are some exceptions, e.g.

- (50) *lafee* [làfee] ‘bone’, *mataa* [màtaa] ‘head’; but *harree* [hárrée] ‘donkey’
 (Cf. Southern: *lafee* [láfèe]; *mataa* [mátàa], [màtaa])

All bound morphemes, except the ones that create extended verbal stems, have their own tonal specification, which contributes to the tonal pattern of the word. For instance, the genitive construction consists of the thing possessed followed by the possessor with lengthening of the last vowel and assignment of high tone, e.g.

- (51) *fira intala kan d’ufte+e* [d’úftée] ‘the friend of the girl who came’
 friend girl that came+GEN
 Cf. *d’ufte* [d’úftè] ‘she came’ (in isolation)

Western Oromo is characterized by displaying a simplified tonal system (cf. Gragg 1976: 173).²⁷ For example, in the Harar Eastern variety, almost all suffixes ending in a long vowel induce a high tone on the syllable preceding them (cf. Owens 1985: 29); however, in Western Oromo they tend to follow the general low–high pattern, though some words escape to this generalization, e.g.

- (52) *gudd+aa* [gúddáa] (Harar Eastern), [gùddáa] (Western) ‘big (masc.)’
gudd+oo [gúddóo] (Harar Eastern), [gùddóo] (Western) ‘big (fem.)’
 Cf. *gog+aa* [gógáa] (Harar Eastern, Western) ‘dry’

Tonal patterns that bear grammatical information cannot vary. Words in variable tone contexts, i.e., in grammatical contexts that do not require a fixed tonal pattern, may vary their basic tonal pattern. Owens (1985: 49) points out that other relatively independent factors may govern variability

27. Significantly, only in this variety some individual words have been found to be distinguished by tone alone, e.g., *gogaa* [gógáa] ‘dry’, *gogaa* [gògáa] ‘skin’. The negative verbal prefix and the focus verbal prefix are also differentiated by means of tone, e.g., *hin* [hín] (focus), [hìn] (negation): *hiraafa* [hìrràfà] ‘he sleeps (focus)’, *hiraifu* [hìrráfù] ‘he does not sleep’.

of tonal patterns in Oromo, namely factors of meaning, according to which certain tonal patterns occur in order to minimize ambiguity, and factors of information, having to do with the distribution of old and new information.

References

- Ali, Mohammed, & Andrzej Zaborski. 1990. *Handbook of the Oromo Language*. Cracow: Polska Akademia Nauk—Oddział W. Krakowie.
- Andrzejewski, B. W. 1957. "Some Preliminary Observations on the Borana Dialect of Galla." *Bulletin of the School of Oriental and African Studies* 19: 354–74.
- Banti, Giorgio. 1988. "Two Cushitic Systems: Somali and Oromo Nouns." In *Autosegmental Studies on Pitch Accent*, ed. H. van der Hulst and N. Smith, pp. 11–49. Dordrecht: Foris.
- Bender, M. Lionel, & Mulugeta Eteffa. 1976. "Galla." In *Language in Ethiopia*, ed. M. L. Bender, J. D. Bowen, R. L. Cooper, & C. A. Ferguson, pp. 130–48. London: Oxford University Press.
- Bryan, Margaret A. 1948. *The Distribution of Semitic and Cushitic Languages of Africa*. London: Oxford University Press.
- Gamkrelidze, Thomas V. 1978. "On the Correlation of Stops and Fricatives in a Phonological System." In *Universals of Human Language*, ed. Joseph H. Greenberg, vol. 2, pp. 9–46. Stanford: Stanford University Press.
- Gragg, Gene B. 1976. "Oromo of Wellegga." In *The Non-Semitic Languages of Ethiopia*, ed. M. L. Bender, pp. 166–95. East Lansing: Michigan State University Press.
- . 1982. *Oromo Dictionary*. East Lansing: Michigan State University, African Studies Center.
- Heine, Bernd. 1981. *The Waata Dialect of Oromo: Grammatical Sketch and Vocabulary*. Berlin: Reimer.
- Heine, Bernd, Thilo C. Schadeberg, & Ekkehard Wolff (eds.). 1981. *Die Sprachen Afrikas*. Hamburg: Helmut Buske.
- Lloret, Maria-Rosa. 1988. "Gemination and Vowel Length in Oromo Morphophonology." Ph.D. dissertation, Indiana University.
- . 1989. "Final Vowels and Grammatical Marking in Oromo." In *Current Approaches to African Linguistics*, vol. 5, ed. Paul Newman and R. B. Botne, pp. 73–83. Dordrecht: Foris.
- Moreno, M. M. 1939. *Grammatica teorico-pratica della lingua galla con esercizi*. Milan: Mondadori.

- Owens, Jonathan. 1980. "Observations on Tone in the Booran Dialect of Oromo." *African Language Studies* 17: 141–96.
- . 1985. *A Grammar of Harar Oromo (Northeastern Ethiopia)*. Hamburg: Buske.
- Sasse, Hans-Jürgen. 1975. "Galla /s/, /š/, und /f/." *Afrika und Übersee* 58: 244–63.
- . 1979. "The Consonants Phonemes of Proto-East-Cushitic (PEC): A First Approximation." *Afroasiatic Linguistics* 7/1.
- . 1981. "Die Kuschitischen Sprachen." In Heine, Schadeberg, & Wolff 1981: 187–215.
- Stroomer, Harry. 1987. *A Comparative Study of Three Southern Oromo Dialects in Kenya: Phonology, Morphology, and Vocabulary*. Hamburg: Buske.

This page intentionally left blank.

Somali Phonology

Annarita Puglielli

Terza Università degli Studi di Roma

26.1. Introduction

Somali belongs to the Cushitic subgroup of the Afroasiatic family. The term Afroasiatic, which was made famous by Greenberg (1955), refers to a large number of languages spoken in northeastern Africa that were previously identified by the term Hamito-Semitic. It includes Semitic, Berber, Egyptian, Chadic, and the controversial Omotic.

The Cushitic branch consists of a large number of languages and can in turn be subdivided into two groups: Highland East Cushitic and Lowland East Cushitic. The first includes such languages as Sidamo, Hadiyya, Kam-bata, and Burji, spoken by about 2 million people in southwestern Ethiopia, along the Rift Valley. Among the major members of the second group are the Oromoid languages (which include Oromo, Konso, Gidole), and the Soma-loid languages. The latter include Rendille, Boni, etc., as well as Somali, which is spoken by about 5 million people in the territory of the Republic of Somalia, the Ogaden (southern part of Ethiopia), Djibouti, and some north-eastern territories of Kenya. There are still some problems of classification for several languages in the Cushitic group. This is not, however, the case of Somali, certainly one of the most well described African languages. Somali has been the object of study since the end of last century and scholars like Berghold (1897, 1899), Schleicher (1892), Reinisch (1903), and Cerulli (1957–59) have contributed to its scientific description (the variance in the description is due to the fact that different authors worked on different regional and/or ethnic varieties of the language).

Although characterized by overall linguistic uniformity, the language that we call Somali does contain dialects, even if they are usually mutually comprehensible ones. We will not go into them here; readers may refer to Cerulli (1957–59), Moreno (1955), and Lamberti (1983). The relevant fact about Somali, though, is that linguistic standardization has been going on for several decades, and that the variety spoken in the Mudug area (central-northern region) has been regularly used since 1943 for radio broadcasting.

This variety was adopted as the national language and officially transcribed in 1972. As for which script to use, a long debate ensued between supporters of Arabic, Latin, and Osmanian (a modified version of the Arabic script created by a Somali scholar, Osman Keenadid). The final decision was in favour of the Latin script. The phonemic notation by Andrzejewski (1955) was chosen for transcriptions, and all special characters or diacritics that had been used in the past were replaced with unused characters in the Latin alphabet (in particular *c* = *ç*, *x* = *ħ*, *dh* = *d̥*).

26.2. The sound pattern of Somali

26.2.1. *Consonants*

Somali has a system of 21 consonant phonemes, each of which has important positional variants (Armstrong 1934, Cardona 1981). This produces a rich and varied consonant system as follows:

Table 26-1. Somali Consonants

	Labial (Dental)	Dental (Alveolar)	Postalveolar	Palatoalveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
stops	b	t d	ɖ			k g	q		ʔ
affrivate				ɖʒ					
nasal	m	n							
lateral		(l)							
rolled		(r)							
fricative	(f)	s		ʃ			χ	ħ ʕ	h
approximant	w					j			

/b/ is almost devoiced in initial position [b̥ʌd] ‘sea’; [β] (affricate without complete closure) between vowels, especially after a stressed syllable [ʔʌβʌ] ‘two’; a voiced stop when geminate in intervocalic position [gabbaɓ] ‘twilight’; devoiced and unreleased [laab̥̚] ‘chest’ or as [pʔ] with glottal closure and release [sapʔ] ‘outcast’ in final position.

/t/ can occur only initially or in intervocalic position; it never occurs in final position, where it neutralizes with /d/. It is strongly aspirated.

/d/ has the same phonetic variation as /b/ with the same distribution.

/q/ is postalveolar, pharyngalized, and voiced; it occurs in initial position (although only in Isaaq, a variety spoken in northern Somalia) as well as in syllable-final position. In common Somali, when in final position, it has become [r]: /baq/ → [bar].

/k/ is aspirated like /t/ and is a voiceless velar; it cannot occur in final position, where it neutralizes with /g/.

/g/ has the same realization as /b/. It can be double in intervocalic position [haggee] ‘where’.

/q/ in initial position is [ḡ] with little voicing; it becomes a voiced fricative in intervocalic position, especially after a stressed syllable. In final position it can be either unexploded and voiceless or more frequently a voiceless fricative uvular.

/dz/ occurs only at the beginning of a syllable and is [ḏʒ] with little voicing; for some speakers it is [tʃ].

/ʔ/ is a glottal stop both in initial and final position. It is often optional in final position but if it is present, it necessarily appears in the presence of suffixes.

/m/ is normally articulated and it occurs only in syllable-initial position. In syllable-final it neutralizes. So we have *nin* ‘man’ whose underlying form is /nim/ so that we have [nin] or [ninka], but [niman] ‘men’.

/n/ appears in all positions.

/l/ its pronunciation varies according to the quality of the vowel that follows it in all positions.

/r/ is [r] in all positions; it can be doubled in middle position [ʕarab] ‘Arabic’ [ʕarrab] ‘tongue’.

Fricatives and approximants are always pronounced according to the description given in the general table and do not show positional variants with the exception of /h/ and /ħ/, which can be voiced in intervocalic position.

26.2.2. *The vowel system*

The vowel system of Somali was first described by Armstrong (1934), who recognized 18 vowel phonemes, and subsequently by Andrzejewski (1955), who described a system of 20 vowels. There is an opposition between “retracting vowels” and “fronting vowels,” and an opposition between short and long vowels. Official orthography does not note the first of these two oppositions, whereas it makes use of two letters (*aa*, etc.) to signal long

vowels. The following is therefore a comprehensive table of the Somali vowel system, with “fronting” or advanced vowels marked with a comma.

Table 26-2. Somali Vowels

Short Vowels			
Fronting		Retracting	
ɪ [i]	ʊ [u]	ɪ [ɪ]	u [ʊ]
ɛ [e]	ɔ [o]	e [ɛ]	o [ɔ]
æ [æ]		a [ʌ, a]	
Long Vowels			
Fronting		Retracting	
ii [i:]	uu [u:]	ii [i:]	uu [u:]
ee [e:]	oo [o:]	ee [ɛ:]	oo [ɔ:]
aa [æ:]		aa [a:]	

One of the peculiarities of the Somali system is that within the same phonological unit only one of the two series of vowels can be present, i.e., there exists a system of vowel harmony. Thus a single word can contain only vowels that belong to the “advanced” set or to the “retracted” one. In addition, this opposition impacts not only on phonology, but also on morphology. In fact, monosyllabic nouns such as *gār* ‘beard’ ~ *gar* ‘will’, *dɔb* ‘skin’ ~ *dub* ‘tail’, *çad* ‘piece’ ~ *cad* ‘white’, which are respectively masculine and feminine, are differentiated on the basis of the vowel quality.

There are also verbs in which the difference between the imperative and the past is marked by fronting vs. retracting vowels: *tɪn* [tɪn] ‘hit it’ ~ *tun* [tʊn] ‘he hit’, *çeg* [eeg] ‘look’ ~ *eeg* [ɛeg] ‘he looked’.

Finally, the interaction of the vowel quality (fronting vs. retracting) and tonal stress—one extremely important feature of Somali grammar—also has consequences on the final consonant sound in a bisyllabic word. The stops (b, d, g) are devoiced in final position (see § 2.1) as in many other languages, but devoicing goes together with releasing or unreleasing according to the position of stress and the quality of the vowel. If the tonal stress is on the last vowel, the stop will be released except where the vowel is fronted. In all other cases—words with stress on the penultimate vowel and/or words with fronted vowels—the stops will not be released (Angoujard & Mohamed 1991).

Somali vowels pose some problems since more parameters are involved in the differentiation of minimal pairs. In addition, only long vowels have a rather constant realization while short vowels vary considerably according to context and different speakers.

26.2.3. *Tonal accent*

The description of this aspect of the phonological system of Somali has been highly discussed and the debate whether Somali is a tone language or not has been going on for decades.

The first descriptions of Somali do not mention tone at all, and some of the grammars mention that accent in Somali falls mainly on the penultimate syllable (Hunter 1880: 6; Reinisch 1903: 28–29). The first to notice the tonal nature of this accent was Armstrong (1934), who distinguishes 4 different tonal levels and notices a strict relationship between tone and accent. In 1949 Klingenheben considers accent to be the primary aspect of Somali and tone to be secondary or derived; thus, on the basis of phonological criteria, he does not consider Somali a tone language but rather something approaching a system with “Starktonsprachen” (‘accent of intensity’) like English or German. More recently Somali has still been considered a tone language characterized by tonal accent without an explicit consideration of the accentual nature of these tonal accents (Jones 1950, Abraham 1964, Andrzejewski 1956).

It is only in recent years that Hyman (1981) proposed a formal description of Somali as a system with tonal accent. From a phonetic point of view Somali shows a contrast between a high tone (A) and a low tone (B). The first has a very limited distribution; each word can have only one A tone and this has to be either on the last vowel or on the penultimate one. The superficial distribution of different surface tonal patterns may be seen in the following examples:

- | | | | | |
|-----|-----------------|----------|-----------------|-------------------|
| (1) | a. <i>ínan</i> | ‘boy’ | b. <i>inán</i> | ‘girl’ |
| | <i>qaálin</i> | ‘little’ | <i>qaalín</i> | ‘little’ (female) |
| | <i>daméer</i> | ‘donkey’ | <i>dameér</i> | ‘donkey’ (female) |
| (2) | a. <i>túug</i> | ‘thief’ | b. <i>tuúg</i> | ‘thieves’ |
| | <i>doofáar</i> | ‘pig’ | <i>doofaár</i> | ‘pigs’ |
| | <i>Soomaáli</i> | ‘Somali’ | <i>Soomaalí</i> | ‘Somalis’ |

The words in (1) and (2) are all nouns. In (1) the tonal alternation in (a) and (b) shows the opposition between masculine nouns (in [a]) and feminine

nouns (in [b]). The masculine nouns are characterized by tone A on the penultimate vowel, while the feminine nouns are characterized by tone A on the last vowel. The alternance shown in (2a) and (2b) is used to mark number. The nouns in (2a) are singular while those in (2b) are plural. One notes that nouns in (2b) are also feminine and their tone A falls on the last vowel as in (1b). The consideration of the long vowel as a unit made up of two units allows us to give a general rule for cases like *ínan/inán* and *daméer/dameér*. The possible generalization is therefore that masculine nouns are characterized by an A tone on the penultimate vowel and feminine nouns by an A tone on the last vowel. As for verbs, the situation parallels that described for nouns (i.e., tone A appears either on the last or on the penultimate vowel). Of course the morphological features relevant here will not be gender and number but rather the mode, the conjugation, or the construction in which the verb occurs (main vs. relative/subordinate clause). Here are some examples of A tone distribution in imperatives and infinitives (the two possible citation forms of verbs) for verbs of different conjugations:

(3)	Conjugation	Imperative	Infinitive
	1	<i>cún</i>	<i>cuní</i>
	2	<i>joóji</i>	<i>joojín</i>
	3	<i>dháqso</i>	<i>dhaqsán</i>

All other word classes fall in one of the two tonal patterns described for nouns and verbs: adjectives are normally considered a separate class of verbs and share the tone patterns of verbs; numerals are nouns and have the same patterns as nouns. Particles normally do not show an A tone. These include prepositions, clitic pronouns for subject and object, impersonal subject pronouns, and focus markers which, from a distributional point of view, all occur in pre-verbal position with fixed relative order.

As shown by the data just examined, an A tone must be present in all nouns, verbs, and other words except particles. It therefore has the same culminative properties we find in languages characterized by stress. A description of the tone in Somali has to explain both its superficial phonetic tonal nature and its underlying accentual identity. Hyman therefore proposes to consider this system, at an abstract level, as an accentual one rather than a tonal one. It is not possible to give here the full description of the whole system and the rules he proposes, but we will try to outline the main points he makes.

First of all accents (and related tone realizations) are not determined by lexical features but rather by grammatical features. Hence underlying forms are registered in the lexicon without prosodic information. Accents are introduced by morphological rules and then, after the application of reduction rules, their tonal realization is specified by means of phonetic rules. The rules for accent assignment are determined by grammatical features. Stress pattern is predictable: for nouns, on the basis of grammatical gender, declension, and type of construction; for verbs, on the basis of modality, conjugation, and type of construction. From a typological point of view, therefore, the system of Somali is of particular interest: it differs from tone systems in that the occurrence of tone A is limited to the last two vowels of a word and only one tone A can occur in a single word. It is also different from languages characterized by an accentual system, since the accent is assigned to a single vowel rather than to the syllable (as in English).

26.3. Phonological rules

Before going into the description of the main phonological rules of Somali, it is important to establish the following facts the relevance of which will be evident in the succeeding description.

First of all, Somali is more accurately described in terms of syllable structure than in terms of morpheme structure. The syllable has the following structure:

$$(4) \quad (C_1)V(V)(C_2)$$

In the underlying form there is no constraint on C_2 , whereas in the surface form the sounds [k, t, d, m] cannot occur syllable final ([d] can occur in this position in *Isaaq*).

Given this syllable structure, a contoid which is phonetically long cannot be considered a single consonant, but rather a sequence of two consonants which are identical but belong to two contiguous syllables. On the basis of the restrictions on the possibility of occurrence of contoids in syllable final position, we deduce that geminates can only be consonants occurring in that position; thus [-tt-, -kk-, -mm-] are not possible. It follows that two consonants can meet only at the boundary of two syllables or at the boundary of two morphemes. When two consonants occur in a single morpheme (for example, in borrowings from Arabic) an epenthetic vowel is introduced,

producing two syllables: *labis*—from Ar. *labs* ‘uniform’, *rodol* from Ar. *ratl* ‘unit of measure’.

An opposite and very general rule of syllable reduction applies instead when we have a sequence of three syllables where the first and second syllable are open and have a short vowel. In this context the middle vowel is generally canceled and we obtain a two-syllable word rather than a three-syllable one. The corresponding rule is:

$$(5) \quad (C)VCVCV \rightarrow (C)VCCV$$

For this rule to apply, the middle vowel must be unstressed. A syllable structure like the one in (5) is often obtained when we add a suffix—either derivational or inflectional—to a root or base, and the rule of syllable reduction applies systematically with very few exceptions (Puglielli 1984a, Puglielli & Ciise 1984). Here are some examples:

(6)	Underlying	Surface	
	<i>orod+ay</i>	<i>orday</i>	‘I ran’
	run+1st past		
	<i>hadal+ay</i>	<i>hadlay</i>	‘I talked’
	talk+1st past		
	<i>fur+am+ayaa</i>	<i>furmayaa</i>	‘it is getting opened’
	open+pass.+3rd progr.		
	<i>hílib+o</i>	<i>hílbó</i>	‘meats’
	meat+plural		
	<i>xárig+o</i>	<i>xargó</i>	‘ropes’
	rope+plural		
	<i>qósol+e</i>	<i>qoslé</i>	‘the one who laughs’
	laugh+agentive		
	<i>qósol+id</i>	<i>qoslíd</i>	‘laugh (noun)’
	laugh+nominalizer		

The second general statement to be made is that most of the rules involving sound modification are grammatically conditioned, i.e. the grammatical status of morphemes determines the sound change they undergo. Consider for example the following rule:

$$(7) \quad /l + t/ \rightarrow /ʃ/ \text{ (sh)}$$

This happens in examples like:

$$(8) \quad qosol + ta \rightarrow qososho \text{ ‘the one who laughs’}$$

where *t* is the initial consonant of the definite feminine article, or the initial consonant of the person agreement suffix in the verb. But in a case like

- (9) *ul+tan* → *ultan* 'stick'

where *-tan* is a derivational morpheme that permits the derivation of a verb from a noun, and therefore has the status of a lexical affix, the sequence /l+t/ does not change into /j/. Another example of the same phenomenon is present in what follows. In words like

- (10) *badda* 'the sea'
qaadday 'you (sing.)/she took (it)'

the noun *bad* plus the feminine definite article *ta* results in *badda* with the assimilation of [t] to [d]. In *qaadday*, to the base *qaad* is added the inflection *-tay* (which is the form of the second singular or the third feminine singular ending for the past tense) which gives *qaadday* where [t] assimilates to the preceding [d]. But in a word like *qaatay* 'I/he took it for myself/himself' (formed by the base *qaad-* plus the root extension for autobenefactive *-at* and the person inflection *-ay*), the result of combining these elements is not the assimilation of the [t] of the lexical extension of the verb to the preceding [d], but rather the simplification of the cluster obtained through the application of the syllable reduction rule described earlier. In other words the stages of the derivation of this word are:

- (11) *qaad+at+ay* → *qaadatay* → *qaadtay* → *qaatay*

Even in this case the fact seems to be due to the lexical nature of the affix *-at*. It seems therefore reasonable to conclude that in general morphophonological rules are grammatically conditioned.

26.3.1. *Juncture phenomena*

In the Somali lexicon, morphemes are mostly monosyllabic but there is a rich system of both verbal and nominal derivation; as a consequence, affixes are added to roots and the final sound of the root joins with the initial sound of the affix. Furthermore, this language has a rich inflectional system both for nouns and for verbs: nouns are marked for gender and number, and verbs for tense/aspect, mood, and person. Finally determiners (i.e., definite articles, demonstratives, possessives, etc.) are all affixed to the noun. Hence juncture phenomena occur in many contexts. We will limit our description to the most common phonological rules determined by this situation for two good reasons: first, because the existing descriptions are still not exhaustive

and there are several aspects that need further investigation, and secondly, because of the introductory nature of this work.

We may start by examining the phenomena connected with the co-occurrence of definite articles with nouns. These articles are respectively *-kV* (where V stands for any of the following: *a, u, ii*) for masculine nouns and *-tV* (*-ta, -tu, -tii*) for feminine nouns. Notice that *-k-* as marker of masculine and *-t-* as marker of feminine is used in other contexts in Somali and is characteristic of other Cushitic languages as well.

When *-kV* is added to masculine nouns we have the following situation:

- (12) */k/* → */g/* when preceded by */g, w, y, i/*
buug+ka → *buugga* 'the book'
rag+ka → *ragga* 'the men'
bari+ka → *bariga* 'the East'
- (13) */k/* → *∅* when preceded by */h, ɸ, ʕ, q, ɣ/*
rah+ka → *raha* 'the frog'
madax+ka → *madaha* 'the head'
- (14) */k/* → */h/* when preceded by a V except */i/*
bare+ka → *baraha* 'the teacher'
aabbe+ka → *aabbaha* 'the father'
biyo+ka → *biyaha* 'the water'

In (12) there is a case of progressive assimilation which is determined by the feature [+back]. In (13) the derivation takes place in two steps: first assimilation and then simplification of the double consonant:

- (15) *rah+ka* → *rah - ha* → */raha/*

The same derivation applies to all the other consonants given in (13) that are all uvular and either fricatives or stops.

In (14) */k/* becomes */h/* when preceded by a vowel except */i/*. This form can only be described synchronically by an ad hoc rule and furthermore, if one looks at the examples, there is more going on than the shift from */k/* to */h/*. In order to explain these forms Cardona (1981: 19–20) hypothesizes that the underlying form is */aabbah/*, */biyah/* to which the following rules apply:

- (16) */aabbah+∅/* → */aabbe/*
/aabbah+ha/ → */aabbah - ha/* → */aabbaha/*

He says that there is no superficial manifestation of this underlying form in today's Somali, but Abraham (1964: 329) makes the assertion that there are

some speakers that pronounce nouns ending in /-o/ and /-a/ as [-oh] and [-ah]; this seems to be confirmed by an instrumental analysis conducted by Farnetani (1981), whose data do not show an abrupt closure of the glottis at the end of the final vowel.

A parallel situation occurs for *-tV* when added to a feminine noun. When preceded by a vowel, /j/, /w/, or /d/, the initial /t/ becomes /d/:

- | | | | |
|------|----------------------|-------------------|----------------|
| (17) | <i>kaneeco+ta</i> → | <i>kaneecada</i> | ‘the mosquito’ |
| | <i>dulmadow+ta</i> → | <i>dulmadowda</i> | ‘the jackal’ |
| | <i>mindī+ta</i> → | <i>mindida</i> | ‘the knife’ |
| | <i>bad+ta</i> → | <i>badda</i> | ‘the sea’ |

/t/ becomes voiced when the preceding noun ends in /q/, /h/, /ɣ/, /ʁ/, /ʔ/. This is obvious for /ʁ/ that is voiced, but not for the other consonants that are voiceless. This fact can be explained in terms of articulatory mechanisms, since a short vocoid seems to be realized at the end of voiceless fricatives by the vibration of the vocal cords, and the same thing happens for /q/ (Cardona 1981: 20).

Finally, when *-tV* follows an /l/ the two sounds become /ʃ/ through assimilation:

- (18) *ul+ta* → *usha* /uʃa/ ‘the stick’

It is rather complex to explain this change: the only proposed explanation is the one by Cardona (1981: 20) based on the fact that /l/ assimilates /n/ in verbal endings and produces /ʃ/ even when followed by /s/ (see *fuul* ‘to climb’ ~ *fuushan* ‘to be on top’), but it can co-occur with /k/ (*wiilkii* ‘the boy’) and /ʃ/ (*bulsho* ‘society’).

The reason for the assimilation of /l+t/ to /ʃ/ has to be attributed to some peculiarity of /l/ when it occurs at the end of a morpheme; it seems to have a fricative component (cf. Armstrong 1934: 126). On these bases Cardona (1981) suggests the following steps:

- (19) [l̥t] → [l̥θ] → [lʃ] → [ʃ]

The last to be taken into consideration is when /t/ follows /d/. Here /t/ assimilates to the preceding sound:

- (20) *gabadh+ta* → *gabadha* [gabaɖɖa]

The following table taken from Cardona (1981) will give an overall picture of the phonological rules that operate between a base ending in a consonant and a suffix beginning with a consonant. The table covers the

majority of the existing cases, since /k, t/ are normally the possible initial consonants of nominal suffixes and /t, s, n/ those for verbal suffixes.

Table 26-3. Somali Sandhi

	k	t	n	s
b	bk	bt		
d	dk	dd		
g	gg	gt		gs
t			nn	
q	q	qd		
ʔ	ʔ	ʔd	nn	
f	fk	ft		
s	sk	st		
ʃ		ʃt		
ħ	ħ	ħd	nn	
χ	χ	χd		
ʕ	ʕ	ʕd		
h	h	hd	hn	
r	rk	rt	rr	
l	lk	ʃ	ll	
n	nk	nt		
j	jg	jd		
w	wg	wd		
V	Vg	Vd		

Before concluding this section we must consider those cases where two vowels meet, one being the last sound of the base and the other the initial sound of the suffix. Here are some examples:

- (21) a. *gudi+id* → *gudiyid* ‘the dangling of the head’
b. *dibi+o* → *dibiyo* ‘bulls’
c. *aabbe+aal*→ *aabbayaal* ‘fathers’
d. *mind+o* → *mindiyoo* ‘knives’

The phonological rule prescribing the insertion of /y/ between two vowels is the same in all cases—whether the root is a verb (as in [21a]) or a noun (as in [21b–d]), and whether the affix is derivational (as in [21a]) or inflectional (as in [21b–d]).

26.3.2. Palatalization

This phenomenon is very general in Somali so that we have for example:

- (22) *jilib* /dʒilib/ ‘knee’

while the initial sound for the same word in other Cushitic languages is /g/ (Bayso *gilib*, Konso-Gidole and Gato *kilba*, Burgi *gilba*, etc.).

In this respect, Somali itself still shows an allomorphic variation /dʒ~ǧ/ as in *joog* ‘to stop’, *jooji* ‘to cause to stop’, *dhergan* ‘to be satisfied’, *dherjin* ‘to satisfy’.

As shown in the example, palatalization occurs frequently in causative verbs where the final consonant of the root is followed by /-i/, part of the causative suffix.

Palatalization occurs also in other cases, like *shimbir* /ʃimbir/, ‘bird’ (Bayso *kimbir*, Afar *kimmiro*), *shan* ‘five’ (Bayso *keni*, Konso *keni*).

26.3.3. Epenthesis

The epenthesis of [n] in front of any consonant—not deriving from dissimilation of a geminate consonant or influence of neighboring sounds—is present in Common Somali and even more frequently in varieties such as Benadir and Digil. Examples are: *ca(n)shuur* ‘tax’, *ma(n)qas* ‘scissors’.

26.3.4. Coalescence

Morphemes and words very often combine so that one obtains a word shorter than its independent parts, as shown in the following examples.

The focus particle *baa*, which necessarily marks an NP, is added to the noun or to the last word of the NP ending in a short vowel or diphthong and we obtain:

- (23) *ninka+baa* = *ninkaa* ‘the man’
maxay+baa = *maxaa* ‘what?’

The presence of the focus particle in words like *ninkaa* is revealed by the long vowel and the tone pattern (Saeed 1987: 25).

When the focus particle *baa* is followed by a clitic subject pronoun, the two will coalesce. See:

(24) *Axmed baan arkay* 'I saw Axmed'

where *baan* is the fusion of *baa* (focus particle) plus *aan* 'subject clitic pronoun 1st person singular'. Coalescence can be even more complex:

(25) *Maxaad doonaysaa?* 'What do you want?'

In (25), *maxaad* is the result of the fusion of three words: *maxay* 'what' + *baa* 'focus' + *aad* 'you'.

26.3.5. Reduplication

This mechanism is widespread in Somali morphology, showing a certain amount of correspondence between form and meaning. All reduplicated forms express the meaning 'more' in terms of intensity or number.

Reduplication can be used both with verbs and nouns: when used with nouns the result is a plural and can be used only with monosyllabic masculine nouns:

(26)	<i>af</i>	<i>afaf</i>	'mouth, tongue'
	<i>nin</i>	<i>niman</i>	'man'
	<i>buug</i>	<i>buugag</i>	'book'

Notice that the plural formation operates on the noun's underlying form (see *nin/niman*).

As for verbs, reduplicated forms have an intensive meaning of repeated action:

(27) *fur* 'to open' *furfur* 'to open more than once'

References

- Abraham, R. C. 1964. *Somali-English Dictionary*. London: University of London Press (repr. 1968).
- Andrzejewski, B. W. 1955. "The Problem of Vowel Representation in the Isaaq Dialect of Somali." *Bulletin of the School of Oriental and African Studies* 17: 567–80.
- . 1956. "Accentual Patterns in the Isaaq Dialect of Somali." *Bulletin of the School of Oriental and African Studies* 18: 103–29.
- Angoujard, Jean, & Mohamed M. Hassan. 1991. "Qualité vocalique, rythme et genre grammatical en Somali." *Linguistique Africaine* 6: 11–49.
- Armstrong, Lilius E. 1934. "The Phonetic Structure of Somali." *Mitteilungen des Seminars für orientalischen Sprachen zu Berlin* 37–38: 116–61 (repr. Westmead Farnborough: Gregg, 1964).

- Berghold, K. 1897. "Somali-Studien." *Zeitschrift für afrikanische und ozeanische Sprachen* 3: 116–98.
- . 1899. "Somali-Studien." *Wiener Zeitschrift für die Kunde des Morgenlandes* 13: 123–98.
- Cardona, Giorgio R. 1981. "Profilo fonologico del Somalo." In Cardona & Agostini 1981: 5–26.
- Cardona, Giorgio R., & Francesco Agostini (eds.). 1981. *Fonologia e Lessico*. Studi Somali 1. Rome: Ministero Affari Esteri.
- Cerulli, Enrico. 1957–59. *Somalia: scritti vari editi e inediti*. Rome: Istituto Poligrafico dello Stato.
- Farnetani, Edda. 1981. "Dai tratti ai parametri: introduzione all'analisi strumentale della lingua somala." in Cardona & Agostini 1981: 27–107.
- Greenberg, Joseph H. 1955. *Studies in African Linguistic Classification*. New Haven: Compass.
- Hunter, F. M. 1880. *A Grammar of the Somali Language*. Bombay.
- Hyman, Larry M. 1981. "L'accento tonale in somalo." In Cardona & Agostini 1981: 109–41.
- Jones, Daniel. 1950. *The Phoneme: Its Nature and Use*. Cambridge: Heffer.
- Klingenheben, A. 1949. "Ist das Somali eine Tonsprache?" *Zeitschrift für Phonetik und allgemeine Sprachwissenschaft* 3: 289–303.
- Lamberti, Marcello. 1983. "The Linguistic Situation in the Somali Democratic Republic." In *Proceedings of the Second International Congress of Somali Studies, University of Hamburg, August 1–6*, 4 vols., ed. T. Labahn, vol. , pp.155–200. Hamburg: Buske.
- Moreno, Martino M. 1955. *Il somalo della Somalia. Grammatica e testi del benadir, darod e dighil*. Rome: Istituto Poligrafico dello Stato.
- Puglielli, Annarita. 1984a. "La derivazione nominale in somalo." In Puglielli 1984b: 1–52.
- (ed.). 1984b. *Aspetti morfologici, lessicali e della focalizzazione*. Studi Somali 5. Rome: MAE.
- Puglielli, Annarita, & Ciise M. Siyaad. 1984. "La flessione del nome." In Puglielli 1984b: 53–112.
- Reinisch, Leo. 1903. *Die Somali-Sprache: Grammatik*. Vienna: Holder.
- Saeed, John H. 1987. *Somali Reference Grammar*. Wheaton, Md.: Dunwoody.
- Schleicher, A. W. 1892. *Die Somali-Sprache*, part 1: *Texte, Lautlehre, Formenlehre und Syntax*. Berlin: Frohlich.

This page intentionally left blank.

Hausa Phonology

Paul Newman

Indiana University

27.1. Introduction

Hausa has the largest number of speakers of any language in sub-Saharan Africa.¹ It is the first language of some 30 million people in northern Nigeria, the Niger Republic, and in scattered communities of settlers and traders in large towns throughout West Africa. In addition, there is also a Hausa-speaking community in the Blue Nile area of the Sudan dating from the beginning of the 20th century.

Hausa is also widely spoken as a second language and is spreading rapidly in its role as a lingua franca. It is extensively used for governmental, educational, and commercial purposes and is employed in the mass media. Hausa language broadcasting, for example, is done not only within Nigeria and Niger, but also by international stations such as the BBC, Voice of America, Radio Deutsche Welle, and Radio Moscow. A number of newspapers appear in Hausa and book publishing is active. Both the Bible and the Koran have been translated into Hausa.

The predominant writing system now employed, which was introduced by the British colonial administration at the beginning of the 20th century, is a modified Roman alphabet, with neither tone nor vowel length represented.² Many Hausas, however, still prefer the Arabic script (called *ajami*), in which the language had been written a century earlier.

Hausa belongs to the Chadic language family, itself a constituent part of the Afroasiatic phylum. In Chadic, Hausa effectively constitutes a group by itself within the West branch. The only other member of the group, Gwandara, which is spoken some 300 kilometers to the south of Hausaland proper, is a creolized offshoot of Hausa rather than a true sister language.

1. This description is part of a Hausa Reference Grammar being prepared with the support of grants from the U.S. Department of Education (PO-17A10037), the National Endowment for the Humanities (RT-21236), and the National Science Foundation (DBS-9107103).

2. An attempt in Niger to promote an orthography in which long vowels were written with double letters proved to be a failure and was eventually dropped.

Lexically, Hausa has borrowed extensively from other languages. Most of its loanwords have come from Arabic, but it has also taken words from Mande, Tuareg, Kanuri, and other neighboring African languages. In this century, vocabulary development has been due primarily to loanwords from English (in Nigeria) and French (in Niger).

Considering the size and geographical extent of Hausa, it exhibits relatively modest dialect variation. The dialect regarded as Standard Hausa is that spoken in Kano State and adjacent areas to the north and south. The dialects spoken to the west in Sokoto and northwest into Niger can be roughly grouped together as Western Hausa.

Some of the most important sources of information on Hausa phonology are Abraham (1959), Greenberg (1941), Klingenberg (1927/28), and Parsons (1970). A comprehensive bibliography of works on the language is Baldi (1977), to which an update has been provided by Awde (1988).

27.2. Phonological inventory

27.2.1. *Consonants*

The consonant phonemes of Standard Hausa (SH) are presented in Table 27-1.

Table 27-1. Hausa Consonants

vl	f	fy	t	c	k	kw	ky
vd	b		d	j	g	gw	gy
gl	ḃ		ḋ	ʼy	ḱ	ḱw	ḱy
vl			s	sh			
vd			z	(j)			
gl			ts				
	m		n				
			l				
			r				
			ṛ				
				y	w	h	ʔ

The letters *c* and *j* represent the affricates [tʃ] and [dʒ] respectively. (In the Hausa of Niger, /j/ is usually pronounced [ʒ].) The /f/ phoneme is variably

pronounced as [p], [f], or [ɸ]. In Western Hausa (WH), it is usually [hw] before /a(a)/ and [h] before other vowels, e.g. WH *hwaadì* ‘fall’ (= SH *faadì*), WH *tàhi* ‘go’ (= SH *tàfi*). Before back rounded vowels it is often pronounced (and, if so, written) as [h], e.g. *dafàa* ‘cook’, *dàfuwaa* = *dàhuwaa* ‘cooking’. The glottalized series includes both laryngealized, implosive stops, indicated by the “hooked” letters *ɓ* and *ɗ*, and glottalized ejectives, indicated by the hooked letter symbol *ƙ* and the digraph *ts*. The non-glottalized counterparts of these consonants are /b/, /d/, /k/, and /s/, respectively. In Katsina and other WH dialects, there is also a palatal ejective /c/, which contrasts with /ts/ before /a(a)/, e.g. WH *càaki* ‘chicks’ vs. *tsàaki* ‘clicking sound in the throat’, both of which are pronounced /tsàaki/ in SH. The glottalized approximant /y/, which occurs in only a few very high frequency words, is a historically recent phoneme, having developed from the sequence /ɗiy-/ via /ɗy/, cf. SH *yaa* with WH *ɗìyaa* ‘daughter’.

The palatalized and labialized velars contrast with their plain counterparts before the vowel /a(a)/, e.g. *gàdaa* ‘duiker’, *gwàdaa* ‘test!’, *gyàɗaa* ‘peanuts’. Before the back/rounded vowels, the velars are all redundantly labialized and before front vowels they are automatically palatalized. These features are not shown in standard orthography, i.e. *doogoo* ‘tall’ = [doo-gwool], cf. the pl. *doogwàayee*; *geefèe* ‘side’ = [gyeefèe], cf. the pl. *gyàffaa*. The exceptional words with a labialized velar followed by /i(i)/, e.g. *gwiibàa* ‘sediment’, *kwiikwiyò* ‘puppy’, are due to historically recent changes of /u/ to /i/, especially in connection with the monophthongization of the */ui/ diphthong to /ii/, i.e. *gwiibàa* < **gwuiɓàa*; *kwiikwiyò* < **kwuikwuyò*.

The palatalized labial /fy/ is lexically infrequent and is often replaced by its plain counterpart, e.g. *fyaacèe* = *faacèe* ‘blow one’s nose’. Some WH dialects—the exact distribution is not clear—also have a set of labialized alveolars as part of their consonant inventory, e.g. WH *twàarii* = SH *tàarii* ‘cough’; WH *ɗwai* = SH *ɗòoyii* ‘stench’.

In word final position /n/ is pronounced [ŋ], e.g. *cân* ‘there’ = [cân]. Speakers of Standard Hausa also commonly pronounce final /m/ as [ŋ], thereby resulting in a merger of the two phonemic nasals, e.g. *maalàm* ‘teacher’ = [maalàm] or [maalàn] .

The symbol *ɾ* is used to distinguish the apical tap or roll from the retroflex flap *r* with which it contrasts, e.g. *ɾahàa* ‘pleasant chatting’, *raanii* ‘dry season’, *màɾgaa* ‘a cassia tree’, *sarkii* ‘emir’. In word-final position, only *ɾ* occurs, e.g. *àshâɾ* ‘obscene language’, *teebùɾ* ‘table’, *baɾ* pre-object form of *bari* ‘leave, let’. The difference between the two rhotics is not indicated in

orthography. In syllable-final position, many speakers, especially in WH dialects, commonly substitute /l/ for either or both R's. The flap is the native Hausa R; the roll has come in through loanwords, primarily from Arabic, Kanuri, and English, from the phonemicization of expressive pronunciation used with ideophones and intensive forms, e.g. *řagařgàzaa* 'shatter', *řamas* 'emphasizes dryness', and from the rhotacization of alveolar obstruents in syllable-final position, e.g. *fařkaa* 'wake up' (< **fadkàa*), cf. *fàrkaa* 'paramour'.

Glottal stop and /h/ are also historically recent phonemes in Hausa, having developed in a similar manner. They both probably existed in the language for some time, in the case of ['] as a phonetic marker of vowel-initial words and, with short vowels, of prepausal position, and in the case of [h] as an allophone of /f/ as well as also being an alternative means of attack for vowel-initial words. Their phonemicization was due to a combination of language-internal sound changes reinforced and/or stimulated by their introduction in medial and initial position in Arabic loanwords, e.g. *řaddu'aa* 'prayer', *sàbà'in* 'seventy', *řazùmmi* 'fasting', *haa'jàa* 'merchandise', *hàmsin* 'fifty', *jaah'ilii* 'ignorance', *shàhaadàa* 'martyrdom'.

The semivowels /y/ and /w/ only occur in syllable onset position. If they are shifted to the coda because of vowel apocope or morphological processes, they automatically alternate with their corresponding vowels, /i/ and /u/, e.g. *màraayaa* 'orphan', *màrain'iyaa* 'female orphan'; *maay'èe* 'sorcerer', *màitaa* 'sorcery'; *sàyi* 'buy' = *sai* (optional clipped form); *bàraawòo* 'thief', *bàraun'iyaa* 'female thief', *baaw'aa* 'slave', *bàutaa* 'slavery', *awoo* 'weighing', *aun'aa* 'to weigh'.

All Hausa consonants can be geminated. (With consonants indicated by a diagraph, only the first letter is written doubled, e.g. /gàsàššee/ 'roasted' is indicated here and in standard orthography as *gàsasshee*.) At an analytical level, geminates can be viewed as a sequence of identical consonants abutting across a syllable boundary, i.e. *baccii* 'sleep' has the canonical form C₁VC₂C₃VV, where C₂ and C₃ happen to be identical. In underived words, only geminate nasals and liquids are common, e.g. *dann'èe* 'suppress', *han-nuu* 'hand' (cf. *hanuu* 'frankincense tree'), *tallee* 'soup pot'; but others do occur sporadically in native words, e.g. *tukkuu* 'bird's crop' and more frequently in loanwords from Arabic, e.g. *hajji* 'the Hajj', *jabb'aa* 'sleeveless robe'. In morphologically derived forms, however, geminates are extremely common, e.g. *râssaa* 'branches', pl. of *reesh'èe*, *hah'hau* 'mount many or

often', pluractional³ of *hau*; *fàffaadaa* 'broad', derived adjective from *faadii* 'breadth', and *zàaɓaɓɓee* 'chosen, adjectival past participle of *zàaɓaa* 'choose'.

27.2.2. Vowels

The vowel phonemes (which are the same for all dialects) are presented in Table 27-2.

Table 27-2. Hauſa Vowels

Short			Long		Diphthongs	
i		u	ii		uu	
e	o		ee	oo		
a			aa		ai	au

Hauſa has five basic vowels, all of which have long and short counterparts. The long vowels have typical IPA values, while the corresponding short vowels are more lax and centralized. In prepausal position, the qualitative difference between the vowels is less pronounced, but the short vowels are easily recognizable as such because they are automatically checked by a glottal closure. The length contrast is only found in open syllables: in closed syllables all vowels are short.

In word-medial position, vowel length functions lexically, e.g. *fiitòò* 'whistling' vs. *fi tòò* 'ferrying'; *faasàa* 'postpone' vs. *fasàa* 'smash'; *duukàa* 'beating' vs. *dukà* 'all'. In final position, however, its function is to a great extent morphological and grammatical, e.g. *hannuu* 'hand' vs. *à hannu* 'in the hand'; *fi ta a* 'going out' vs. *fi ta* 'go out'; *shi* 'him (direct object form)' vs. *shii* 'him (independent form)'; *saaboo* 'new' vs. *Saabo* 'proper name'.

The present balanced vowel system derives historically from a skewed system in which the number of contrasts varied depending on the position within the word. In final position, all five vowels occurred, but with a minimal length contrast, if at all. The rule seems to have been that apart from monosyllabic content words (nouns and verbs) ending in /aa/, all final vowels were

3. The term "pluractional" refers to a verbal derivation, traditionally called "intensive" in Hauſa and other languages, which is used to indicate the plurality of the action and/or of the patient (either object of a transitive verb or subject of an intransitive), see Newman (1990).

short. In word-initial position—and Hausa did have vowel-initial words—only short /a/ and /i/ were used. (If [u] occurred, it would have been a conditioned variant of /i/.) Otherwise, the language had three vowels (/i(i)/, /a(a)/, and /u(u)/), which could occur long or short.

The vowels /i/ and /u/ exhibit special restrictions in relation to their semi-vowel counterparts. The first is the fact that /i/ before /y/ and /u/ before /w/ are always short.⁴ The second is the fact that whereas the sequences /yi/ and /wu/ occur, */yu/ and */wi/ normally do not.⁵ (Surface exceptions result from the fact that /u/ before /y/ is pronounced [i], i.e. *wuyàa* ‘neck’ → [wiyàa].)

The vowels /e(e)/ and /o(o)/, which underlyingly are always long in medial position, only occur in open syllables. If the syllable becomes closed, due to any number of morphophonological processes, the vowel invariably shortens and generally merges with short /a/, e.g. *gyeefèe* ‘side’, pl. *gyâffaa* (< **gyêffaa* < **gyêeffaa* < **gyeefâfaa*), *toonàa* ‘dig up’, pluractional *tan-tònaa* (< **tontònaa* < **toontònaa*).

Synchronically there are two diphthongs /ai/ and /au/, which function as long vocalic nuclei, e.g. *mâi* ‘oil’, *kaifii* ‘sharpness’, *sàu* ‘times’, *baunaa* ‘buf-falo’. In the not so distant past, there were two other diphthongs: */ui/, which monophthongized to /ii/, e.g. **gwuibàa* > *gwiiàa* ‘sediment’, and */iu/, which monophthongized to /uu/, e.g. **shiukàa* (< **shibkàa*) > *shuukàa* ‘to sow’. (Note the instructive doublet meaning ‘illness’: *cùutaa* (< **cùutaa* < **ciw-taa* (with the *-taa* abstract suffix) = *ciwòo*.) The /ai/ diphthong is pronounced [ai] or [əi] when occurring in a monosyllabic word with falling tone or when preceded by /ʔ/ or /h/. Elsewhere it is pronounced as [ei] or often even [ee], thereby merging with long /ee/. The back diphthong /au/ varies in the [au] to [ou] range, but usually remains distinct from /oo/.

27.2.3. Tone

Hausa has two level tones, Hi (unmarked in transcription) and Lo (indicated by a grave accent), e.g. *raanaa* ‘sun, day’, *dàgà* ‘from’, *taagàa* ‘window’, *bàara* ‘last year’, *gooràa* ‘bamboo’, *gòoraa* ‘large gourd’. (With long vowels, the tone mark is only placed on the first of the two vowels.) It also has a Falling contour (indicated by a circumflex), which only occurs on heavy syl-

4. As pointed out by Gouffé (1965: 195n), the usually reliable dictionary by Abraham (1962) is systematically wrong in this regard: Abraham invariably transcribes long /ii/ and /uu/ before /y/ and /w/ respectively.

5. This asymmetry is contrary to the supposed norm, discussed by Ohala and Kawasaki (1984: 122–24), which is that /yi/ and /wu/ lack acoustic salience and thus are marked in relation to /yu/ and /wi/ and would be expected to occur less commonly than the latter pair.

lables. This Falling tone can be analyzed as Hi plus Lo on a single syllable, e.g. *yâaraa* (HiLo-Hi) ‘children’, *mântaa* (HiLo-Hi) ‘forget’, *mîn* (HiLo) = *minî* ‘to me’.⁶ There is no Rising tone. Presumed Lo-Hi sequences on a single syllable, which would be expected to result from apocopation and such, do not surface. They are simplified to Lo if immediately preceded by Hi, and Hi elsewhere, e.g. *gawâyîi* = *gawâi* (< **gâwâi*) ‘charcoal’, *mukà yî* = *mukâi* (< **múkâi*) ‘we did’; *tàawa* = *tau* (< **tâû*) ‘mine’; *tàusâyîi* = *tàusai* (< **tàusâi*) ‘pity’.

27.3. Syllable structure

Only three syllable types occur in the language: CV, CVV (where VV can be a long vowel or a diphthong), and CVC. (In a few instances, a syllabic nasal serves as the vocalic nucleus of a syllable or as a syllable by itself, e.g. *ngùlu* or *ngùlu* = *ùngùlu* ‘vulture’; *ñ zoo?* or *ñ zoo?* = *’ñ zoo?* ‘Should I come?’) The CV syllable type is light; the other two are heavy. Syllables may not contain both a long vowel (whether monophthongal or diphthongal) and a final consonant. Such overheavy syllables, which commonly result from morphological processes, are automatically pared down by nucleus reduction rules, e.g. **râi-n-sà* (lit. life-of-him) → *rânsà* ‘his life’; **faar-koo* → *farkoo* ‘beginning’ (cf. *faarâa* ‘begin’); *cuus-cùusaa* → *cuccùusaa* ‘stuff repeatedly’.

All Hausa syllables (and thus all Hausa words) begin with a consonant. Words that appear in the orthography with an initial vowel begin phonemically with a glottal stop, e.g. *aure* ‘marriage’ = *’auree/*. This restriction against vowel-initial words is *not* an inherited Afroasiatic feature as it might first appear. Rather it is due to a historically shallow change whereby a prothetic, originally sub-phonemic, consonant, */’/* or */h/*, was added to vowel-initial words, e.g. **askîi* > *’askîi* ‘shaving’, **abâa* > *habâa* ‘chin’. True consonant clusters are not allowed, although two consonants may abut across a syllable boundary, e.g. *han.tâa* ‘liver’. Most words end in a vowel, the exceptions being ideophones, e.g. *wulik* ‘emphasizing blackness’, recent loanwords, e.g. *kyât* ‘cake’, or the result of vowel apocopation, e.g. *kâĩ* = *kadà* ‘don’t’.

Hausa words tend to be disyllabic, trisyllabic, or even quadrisyllabic. Monosyllabic words occur, but with a more restricted distribution. They are the norm for pronouns, connectors, and other function words and they are

6. Although structurally the Falling tone is easily analyzed as a combination of Hi + Lo, there is evidence that at the level of *Sprachgefühl* it exists as a unitary contour.

also common with ideophones. On the other hand, there are only some twelve monosyllabic verbs, mostly *Ci* or *Caa* with Hi tone (e.g. *bi* ‘follow’, *jaa* ‘pull’) and a small number of CVV nouns with Hi or Falling tone, e.g. *faa* ‘flat rock’, *mâi* ‘oil’, *sau* ‘foot’.

27.4. Phonotactic restrictions

27.4.1. Sonorants

In normal CVCV sequences, /l/ and /n/, and /l/ and /r/ cannot co-occur. (As a result, the rendition of the English loanwords ‘linen’, /liɪn/ and ‘nylon’ /liɪlɪn/ end up almost being identical.)⁷ The restriction does not, however, apply to the plural suffix *-unàa*, e.g. *tùuluu* ‘waterpot’, pl. *tuulunàa*, or to other cliticized elements, e.g. *tùuluunaa* ‘my waterpot’. The /r/ restriction only applies to the flap /r/; words with the sequence /r/ – /l/ and /l/ – /r/ do occur, e.g. *lùurà* ‘look after’, *ṛuulàa* ‘ruler’. In the case of flap /r/ and /n/, there is a one-directional restriction: /r/ – /n/ occurs readily, e.g. *rinàa* ‘dye’, *raanii* ‘dry season’, but /n/ – /r/ does not, the word *narkèe* ‘melt’ being an exception.

27.4.2. Glottalized segments

There are two restrictions that affect the glottalized consonants.

(a) One cannot have two different glottalized consonants in the same word, i.e. /b/ – /ts/ or /k/ – /d/ do not co-occur in the same word, whether in immediate sequence or separated by other elements. The Arabic loanword *dâriikàa* ‘religious sect’ is an exception. One can, however, have multiple instances of the same consonant, e.g. *ḡaabèe* ‘quarrel’, *dàadumàa* ‘drive away’, *tsaatsàa* ‘rust’, *kuukùutaa* ‘try hard’.

(b) Generally speaking, glottalized consonants and their non-glottalized counterparts cannot co-occur in the same word, i.e. sequences of /b/ – /ḡ/ or /k/ – /ḡ/ do not occur in either order. On the other hand, whereas the sequence /d/ – /ḡ/ does not occur in that order, /d/ – /d/ is quite normal, e.g. *daadii* ‘pleasantness’, *dadèe* ‘last long’. Similarly, /ts/ – /s/ does not occur, although there are a few examples of /s/ – /ts/ with an intervening consonant, e.g. *santsii*, ‘slipperiness’, *sartsèe* ‘splinter’. Instances of words containing /ḡ/ and a suffix with /k/, e.g. *kauy-ukàa* ‘villages’, *kar-koo* ‘durability’, suggest that the glottalization restriction is perhaps a property of roots rather than full words.

7. Syllable-final nasals have an ambiguous status in Hausa (behaving in many respects like components of the vocalic nucleus) and thus would not be expected to obey the same principles.

27.4.3. Tone and vowel length

With some exceptions (e.g. recent loanwords and ideophonic reduplicatives) there are no words ending in a Lo-Lo tone sequence and a long final vowel. Thus, *màcè* ‘woman’, *gwàdò* ‘blanket’, and *kaařùwà* ‘prostitute’ occur, but words such as **kàrèè* or **zòomòò* or **tunkiyàa* do not.⁸

27.5. Syllable weight

Syllable weight plays an essential role in Hauša phonology and morphology in a number of different areas, of which the following are only some selected examples (see Newman 1972, 1981).

27.5.1. Canonical shape

(a) Different pronoun paradigms are generally marked by a fixed weight pattern. For example, direct object and subjunctive subject pronouns are all characterized by a light syllable, whereas disjunctive and perfective pronouns are characterized by a heavy syllable, e.g. (‘1, 2m., 2f., 3m., 3f., 1pl., 2pl., 3pl.’):

direct object (Hi tone set): *nì, kà, kì, shì, tà, mù, kù, sù*

subjunctive: *‘nì, kà, kì, yà, tà, mù, kù, sù*

disjunctive: *nii, kai, kee, shii, ‘ita, muu, kuu, suu*⁹

perfective: *naa, kaa, kin, yaa, taa, mun, kun, sun*

(b) Abstract nouns of sensory quality, a group of semantically related nouns ending in *-ii*, all have a heavy first syllable, e.g. *zaafii* ‘heat’, *nauyii* ‘heaviness’, *karfii* ‘strength’.

(c) Verb + noun compounds in which the first element is monosyllabic invariably have a heavy first syllable, even if the verb would normally have a short vowel, e.g. *shàa-zumaamì* ‘sugar ant’ (lit. drink honey); *cii-raani* ‘dry season work’ (lit. eat dry season < *ci* ‘to eat’); *bii-bango* ‘water dripping along the wall’ (lit. follow the wall < *bi* ‘to follow’); *kàs-dafi* ‘a poison antidote’ (lit. kill poison).

27.5.2. Rhythmic weight polarity

(a) With the verbalizing suffix *-a(a)ta*, the length of the /a(a)/ is determined by the weight of the preceding syllable to produce either a Heavy-Light

8. This phonotactic restriction was first pointed out explicitly by Leben (1971). For a reinterpretation of various synchronic tone rules postulated in connection with this restriction, see Newman and Jaggard (1989).

9. The feminine pronoun *‘ita* illustrates the metrical equivalence of a heavy syllable to two light syllables.

pattern or a Light–Heavy alternation, e.g. *tsooràtaa* ‘frighten’ < *tsòoroo* ‘fear’; *kàunatàa* ‘love s.o.’ < *kàunaa* ‘love’; *d’anyàtaa* ‘moisten’ < *d’anyee* ‘fresh, moist’; cf. *fùsaatà* ‘be angry’ < *fushii* ‘anger’; *wàdaatàa* ‘enrich’ < *wàdaa* ‘wealth’.

(b) The reduplicative verbalizer also exhibits weight polarity, e.g. *zaafâfaa* ‘to make hot’ < *zaafii* ‘heat’, *kaifâfaa* ‘sharpen’ < *kaifii* ‘sharpness’; cf. *dumàamaa* ‘warm up’ < *dùmii* ‘warmth’.

27.5.3. Syllable weight and tone

(a) Basic disyllabic intransitive verbs ending in *-a* (the so-called “grade 3” verbs) normally have Lo-Hi tone and a light first syllable, e.g. *tùma* ‘jump’, *shìga* ‘enter’, *tsìra* ‘germinate’. Grade 3 verbs with a heavy first syllable generally have Hi-Hi tone, e.g. *girma* ‘grow up’, *kaura* ‘migrate’, *tsiira* ‘escape’.

(b) Plurals of “ethnonyms” (which includes occupations and such) are formed by means of a suffix *-aawaa*. Those which are built on disyllabic stems with a heavy first syllable often have a Lo-Lo-Hi tone pattern; ethnonyms with all other syllabic shapes have all Hi tones, e.g. *Hàusàawaa* ‘Hausa people’, *Gwàaràawaa* ‘Gwari people’, *dùukàawaa* ‘leather workers’; cf. *Badaawaa* ‘Bade people’, *Kanaawaa* ‘people from Kano’, *Zazzagaawaa* ‘people from Zaria’.

(c) Hypocoristics formed by final syllable reduplication have Lo-Hi-Hi tone if the initial syllable is heavy. If the initial syllable is light, the word has an initial Hi tone, the other tones being unpredictable, e.g. *Làadiidi* < *Laadi* ‘fem. name’; *Àduudu* < *Audu* ‘masc. name’; *Mùd’dēede* < *Mùd’de* ‘masc. name’; cf. *Inuunu* < *Inuu* ‘masc. name’, *Kulùulu* < *Kulù* ‘fem. name’.

27.6. Functioning and operation of tone

27.6.1. Lexical and grammatical tone

Tone functions both lexically and grammatically. Although lexically tone does not have a functional load comparable to a West African language such as Igbo, it does serve to distinguish a number of lexical items from one another, e.g. *kai* ‘you (m.)’, *kâi* ‘head’; *gòoraa* ‘large gourd’, *gooràa* ‘bamboo’; *kuukàa* ‘baobab tree’, *kuukaa* ‘crying’; *wuyàa* ‘neck’, *wùyaa* ‘trouble’; *suntàa* ‘untie, loosen’, *sùntaa* ‘catch fish’. Grammatically, tone serves to mark tense/aspects, verb derivations, nominalizations, adverbializations, etc. This is sometimes done by tone itself but more often in conjunction with changes in vowel length, e.g.

taa ‘she (past tense)’, *tāa* ‘she (future tense)’
dafāa ‘to cook’, *dāfaa* ‘cook! (imperative)’
shaa ‘to drink’, *shāa* ‘drinking’; *ci* ‘to eat’, *cīi* ‘eating’
’idòo ‘eye’, *’ido* ‘in the eye’; *kasaa* ‘ground’, *kasà* ‘on the ground’

27.6.2. Fixed tone patterns

Inflectional and derivational constructions in Hausa tend to have associated with them set tone patterns which override the lexical tone of the underlying items. For example, agentives with the prefix *ma-* have the tone pattern Hi-(Lo-)Lo-Hi; language names and locative nouns with the prefix *ma-* are all Hi; abstracts with the ending *-(n)takaa* have the pattern Lo-Lo-Hi-Lo; plurals with the suffix *-unaa* have a ...Hi-Lo pattern; and plurals with the suffix *-ai* have a ...Lo-Hi pattern.

mandómii ‘farmer’, *maròowàcìi* ‘miser’, *majèemii* ‘tanner’
Laañabcii ‘Arabic’, *Jaamusancii* ‘German’
makañantaa ‘school’, *majeemaa* ‘tannery’
jàaṛùntakàa ‘bravery’, *yàaràntakàa* ‘youthfulness’
tùddai ‘hills’, *bàalìgai* ‘adults’, *’àlmùbàzzàṛai* ‘spendthrifts’
huununàa ‘caps’, *agoogunàa* ‘clocks’; *bakunkunàa* ‘bows’

27.7. Phonological processes

Understanding of Hausa phonology and morphology requires the recognition of a number of phonological processes. Some of these processes are essentially historical in nature, although their existence is still evident in morphological alternations; whereas others still function as synchronic rules, although not necessarily in a totally productive, exceptionless manner. For convenience, one can divide the changes into two groups, those affecting coda (i.e. syllable-final) consonants,¹⁰ and those showing the influence of vowels on syllable initial consonants.

27.7.1. Assimilation of nasals

In prevocalic position, there are two distinct nasals, the bilabial /m/ and the alveolar /n/. When followed immediately by another consonant, whether within the same word or across word boundaries, /n/ always undergoes anticipatory assimilation to the position of that abutting consonant, e.g.

10. A number of these changes were described over a half century ago by Klingenberg (1927/28). Some of these, especially the weakening of velar and labial stops to /u/, have now come to be called “Klingenberg’s laws.”

sun bi ‘they followed’ → [sumbi]

gidankù ‘your house’ → [gidaŋkù] (cf. [gidansù]/[gidammù] ‘their/our house’)

In Standard Hausa, but not in WH dialects, /m/ codas also undergo place assimilation, e.g.

dìnkàa [dĩŋkàa] ‘sew’ = WH *dumkàa*

kàzântaa ‘filthiness’, cf. *kàzaamii* ‘filth’.

27.7.2. Syllable-final velars

Syllable-final velars historically weakened to /u/, e.g.

talauçìi ‘poverty’ < **talakçìi*, cf. *talàkà* ‘common man’

sàraunìyaa ‘queen’ < **sàraknìyaa*, cf. *sarkii* ‘king’

haurèe ‘tooth’ < **hakrèe*, cf. *hakoorii* ‘tooth/teeth’ (originally a plural form)

27.7.3. Syllable-final labials

Syllable-final labial obstruents historically weakened to /u/, but in Standard Hausa only. The change also affected /m/, but only when abutting with a following /n/ or /r/. In WH dialects, the original consonant remained.

sàuka ‘get down’ = WH *sàpka*

Audù ‘proper name’ = WH *Abdù*¹¹

’audùgaa ‘cotton’ = WH *’abdùgaa*

shuukàa ‘to sow’ (< **shiukàa*) = WH *shipkàa*

taushii ‘type of drum’, (pl. *tafàashee*) = WH *tafshii*

zaunàa ‘sit’ (cf. *zamaa* ‘sitting’) = WH *zamnàa*

’auree ‘marriage’ (cf. *’amaryaa* ‘bride’) = WH *’amree*

27.7.4. Rhotacization

Syllable-final coronal stops and the ejective /ts/ undergo rhotacization to the rolled /r/ as an active process.

kaĩkàdaa ‘beat repeatedly’ < **kadkàdaa*, cf. *kadàa* ‘beat’

maĩmàtsaa ‘push, pester repeatedly’, cf. *matsàa* ‘push, pester’

hũrhudu ‘four each’, cf. *hudu* ‘four’

’yaĩsà = *’yaatasà* ‘his daughter’

faĩkee ‘trader’ (< **fatkee*), cf. pl. *fatàakee*

kãr apocopated form of *kadà* ‘do not!’

11. Interestingly, names such as *Àbdùllaahì*, *Àbdùlsàlàam*, and *Àbdùlmaalik* are pronounced with /b/ rather than /u/ even by Standard Hausa speakers who say *Audù* rather than *Abdù*.

The rule also applies to /s/ and /z/, but in a more sporadic, unpredictable fashion, e.g.

marār = *maràs* ‘lacking’, cf. pl. *maràsaa*
fitaṛ = *fitas* (WH) ‘take out’ (from *fita* ‘go out’)
gĩrgijèè (< **gizgizèè*) ‘raincloud’, cf. pl. *gìzàagìzai*
mārmaza ‘very quickly’, cf. *maza* ‘quickly’
 cf. *kaskoo* ‘bowl’ (not **kaṛkoo*); *fizgaa* ‘grab’ (not **fĩrgaa*)

27.7.5. Gemination

Coda consonants in Haua commonly assimilate fully to an abutting consonant to form a geminate, e.g.

**fīt-shee* → *fīsshee* ‘take out (pre-pronoun form)’
**kwan laafiyàa* → *kwal laafiyàa* ‘rest well!’
fuskàrkà = *fuskàkkà* ‘your face’

The gemination particularly shows up in reduplicative constructions such as pluractional verbs or adjectives derived from sensory quality nouns. In these cases, gemination occurs instead of the rules changing velars and bilabials to /u/ and alveolars to /ṛ/, the latter usually being a possible option.

daddàkaa pluractional of *dakàa* ‘pound’ (not **daudàkaa*)
kakkàfaa pluractional of *kafàa* ‘affix’ (not **kaukàfaa*)
fīfīta (= *fīṛfīta*) pluractional of *fita* ‘go out’
mammàtsaa (= *mārmàtsaa*) pluractional of *matsàa* ‘push, pester’
kakkaawoo pluractional of *kaawoo* ‘bring’
gwàggwaaḃaa ‘thick’ < *gwaaḃii* ‘thickness’
zàzzaafaa ‘hot’ < *zaafii* ‘heat’
wàwwaaraa ‘smelly’ < *waarii* ‘stench’
kàkkauraa ‘stout’ < *kaurii* ‘stoutness’

27.7.6. Palatalization

(a) When followed by a front vowel, either *i(i)* or *e(e)*, the alveolars *s*, *z*, and *t* palatalize to *sh*, *j*, and *c*, respectively, e.g.

gàshe stative form of *gasàa* ‘roast’
kàajii pl. of *kàazaa* ‘hen’
sàacee pre-pronoun form of *sàataa* ‘steal’

The voiced stop *d* also palatalizes to *j*, with resultant neutralization of the *z/d* contrast, but less regularly than with the above consonants. It generally is blocked by a preceding /n/ and in other environments the palatalization is either optional or lexically determined, e.g.

gìndii ‘base’, *kundii* ‘pad of paper’, cf. *hanjìi* (< **hanzìi*) ‘intestines’

jìdee pre-pronoun form of *jìdaa* ‘transport’

bìdoodii pl. of *bìdàa* ‘thatching needle’

kadoodii = *kadoojii* pl. of *kadàa* ‘crocodile’

gàajee (not **gàadee*) pre-pronoun form of *gàadaa* ‘inherit’

The sonorants *n*, *l*, *r*, and *ɾ* do not undergo palatalization, nor does *d*. The ejective sibilant *ts* palatalizes to *c*’ (an ejective affricate) in WH dialects, but not in Standard Hausa, e.g. SH *duutsèe* = WH *duuc*’i ‘stone’, SH *tsiilaa* = WH *c’iilaa* ‘tapeworm’.

The palatalization process shows up readily in morphological constructions. Because of the introduction of loanwords and the operation of various sound changes, non-palatalized alveolars before front vowels are now very common, e.g.

sillee (= *sullee*) ‘top of corn stalk’

koosee (= *koosai*) ‘fried beancake’

zîi ‘diamonds’ (card suit), *ziinaaɾii* ‘gold’

teebûr ‘table’, *’asìbiti* ‘hospital’ (pl. *’asibitoocii*, with palatalization!)

Because the palatalization rule was presumably fully regular at an earlier historical stage, one now finds words with the palatalized consonant in the underlying, lexical form and the non-palatalized counterpart in the derived form, e.g.

gaashii ‘hair’, *gãrgaasaa* (< **gàsgaasaa*) ‘hairy’

dùkushii ‘colt’, *dùkusaa* ‘female colt’

mijii ‘husband’, pl. *mazaa* ‘males’

kuncii ‘restricted’, *kuntàtaa* ‘restrict’

As was mentioned earlier, /e/ normally changes to /a/ in closed syllables. It should be pointed out that the application of the rules must be ordered so that the palatalization takes place before the /e/ → /a/ change, e.g.

jeenèe (< **zeenèe*) ‘row of reaped corn’ pl. *jânnaa* (not **zânnaa*)

sàataccân ‘the stolen one’ (< **sàatattee-n*)

(b) The semivowel *w* also palatalizes regularly to *y*, e.g.

kàasuwaaw ‘market’, pl. *kaasuwooyii*

baraawòo ‘thief’, pl. *baraayii*

rawaa ‘dancing’, pl. *raaye-raaye*

Viewed in the context of the palatalization rule affecting alveolar obstruents, the *w* → *y* change seems totally ad hoc, but this is not so. The velars *k*, *g*, *ƙ* (and their labialized counterparts) also undergo palatalization, e.g. *taagàa* ‘window’, pl. [taagwoogyii] (orthographically *tagogî*), *ràkee* [ràkyee], ‘sugarcane’, cf. *ràkensà* [ràkyansà] ‘his sugarcane’. This change, unlike the palatali-

zation of alveolars, is not, however, noted in the orthography and thus it has tended to be neglected by Hausaists. Recognizing it, however, allows one to view the $w \rightarrow y$ rule as part of a more general pattern.

27.7.7. Alternation of *f* and *h*

Much less active synchronically than palatalization is the change of /f/ to /h/ when followed by one of the back rounded vowels /u(u)/ or /o(o)/, e.g. *tàfi* ‘go’, *tahoo* ‘come’; *dafàa* ‘cook’, *dàhuwaa* ‘cooking’. It is evident not only in morphological alternations and dialect variants, but also in the realization of a number of English loanwords, e.g.

hudu = *fudù* ‘four’

mahuucii = *mafiicii* ‘fan’

mahòo ‘a patch’, cf. *mafèe* ‘to patch’

tsoohoo ‘old’, cf. pl. *tsòofàffii*

hòotoo ‘photo’ (< Eng.)

hòoloo ‘polo’ (< Eng.)

The following two examples with “irregular” morpheme alternants illustrate the interaction of the various phonological processes that have been described.

- (i) *màkaahòo* ‘blind person’ (*f > /h/ before /oo/); *màkàafii* ‘pl.’ (set ...Lo-Hi tone pattern; surfacing of the underlying /f/);
màkaunìyaa ‘blind woman’ < **màkaaf-nìyaa* (syllable final *f > u;
shortening of /aa/ to /a/ to avoid an overheavy syllable)
- (ii) *zuuci yaa* ‘heart’ < **zuktì-yaa* (syllable final *k > u; palatalization of /t/); *zukaàtaa* ‘pl.’ (set Hi-Lo-Hi tone pattern, surfacing of underlying /k/ and /t/).

References

- Abraham, R. C. 1959. *Hausa Literature and the Hausa Sound System*. London: University of London Press.
- . 1962. *Dictionary of the Hausa Language*, 2nd ed. London: University of London Press.
- Awde, Nicholas. 1988. “A Hausa Language and Linguistics Bibliography 1976–86 (including supplementary material for other years).” In *Studies in Hausa Language and Linguistics in Honour of F. W. Parsons*, ed. by Graham Furniss and Philip J. Jaggard, pp. 253–78. London: Kegan Paul International.

- Baldi, Sergio. 1977. *Systematic Hausa Bibliography* (Istituto Italo-Africano, Collana di Studi Africani 3). Rome: Tip. Pioda.
- Gouffé, Claude. 1965. "La lexicographie du haoussa et le préalable phonologique." *Journal of African Languages* 4: 191–210.
- Greenberg, Joseph H. 1941. "Some Problems in Hausa Phonology." *Language* 17: 316–23.
- Klingenheben, August. 1927/28. "Die Silbenauslautgesetze des Hausa." *Zeitschrift für Eingeborenen-Sprachen* 18: 272–97.
- Leben, William R. 1971. "The Morphophonemics of Tone in Hausa." In *Papers in African Linguistics*, ed. by C.-W. Kim and Herbert Stahlke, pp. 201–18. Edmonton: Linguistic Research.
- Newman, Paul. 1972. "Syllable Weight as a Phonological Variable." *Studies in African Linguistics* 3: 301–23.
- . 1981. "Syllable Weight and Tone." *Linguistic Inquiry* 12: 670–73.
- . 1990. *Nominal and Verbal Plurality in Chadic* (Publications in African Languages and Linguistics, 12). Dordrecht: Foris.
- Newman, Paul, and Philip J. Jaggar. 1989. "Low Tone Raising in Hausa: A Critical Assessment." *Studies in African Linguistics* 20: 227–51.
- Ohala, John J., and Haruko Kawasaki. 1984. "Prosodic Phonology and Phonetics." *Phonology Yearbook* 1: 113–27.
- Parsons, F. W. 1970. "Is Hausa Really a Chadic Language? Some Problems of Comparative Phonology." *African Language Studies* 11: 272–88.

VOLUME 2

PHONOLOGIES
OF
ASIA
AND
AFRICA

Editor
Alan S. Kaye

Technical Advisor
Peter T. Daniels

Phonologies of Asia and Africa

This page intentionally left blank.

Phonologies of Asia and Africa

(Including the Caucasus)

Volume 2

edited by

Alan S. Kaye

Technical Advisor

Peter T. Daniels

Winona Lake, Indiana

EISENBRAUNS

1997

© Copyright 1997 Eisenbrauns

All rights reserved

Printed in the United States of America.

Library of Congress Cataloging in Publication Data

Phonologies of Asia and Africa : (including the Caucasus) / edited by
Alan S. Kaye ; technical advisor, Peter T. Daniels.

p. cm.

Includes bibliographical references.

ISBN 1-57506-017-5 (vol. 1: cloth : alk. paper). — ISBN 1-57506-018-3
(vol. 2: cloth : alk. paper). — ISBN 1-57506-019-1 (2-vol. set: cloth : alk.
paper)

1. Asia—Languages—Phonology. 2. Africa—Languages—
Phonology. I. Kaye, Alan S. II. Daniels, Peter T.

P381.A75P48 1997

414—dc21

97-4964

CIP

The paper used in this publication meets the minimum requirements of the
American National Standard for Information Sciences—Permanence of Paper
for Printed Library Materials, ANSI Z39.48-1984. ☉

Contents

Volume 1

List of Maps	ix
List of Tables	ix
Preface	xiii
Introduction	xv

1 Afroasiatic Languages

Semitic Languages

Ancient and Medieval

East Semitic

1 <i>Akkadian and Amorite Phonology</i>	3
Giorgio Buccellati	
2 <i>Eblaite Phonology</i>	39
Cyrus H. Gordon	

Central Semitic

3 <i>Ugaritic Phonology</i>	49
Cyrus H. Gordon	

Northwest Semitic

4 <i>Phoenician and Punic Phonology</i>	55
Stanislav Segert	
5 <i>Ancient Hebrew Phonology</i>	65
Gary A. Rendsburg	
6 <i>Tiberian Hebrew Phonology</i>	85
Geoffrey Khan	
7 <i>Jewish Palestinian Aramaic Phonology</i>	103
Geoffrey Khan	
8 <i>Old Aramaic Phonology</i>	115
Stanislav Segert	
9 <i>Classical Syriac Phonology</i>	127
Peter T. Daniels	

vi Contents

- 10 *Modern and Classical Mandaic Phonology* 141
Joseph L. Malone

South Semitic

- 11 *Old South Arabian Phonology* 161
Gene Gragg
- 12 *Ge'ez Phonology* 169
Gene Gragg

Modern

Central Semitic

- 13 *Arabic Phonology* 187
Alan S. Kaye
- 14 *Moroccan Arabic Phonology* 205
Jeffrey Heath
- 15 *Cypriot Arabic Phonology* 219
Alexander Borg
- 16 *Maltese Phonology* 245
Alexander Borg

Northwest Semitic

- 17 *Israeli Hebrew Phonology* 287
Shmuel Bolozky
- 18 *Modern Aramaic Phonology* 313
Robert D. Hoberman

South Semitic

- 19 *La phonologie des langues sudarabiques modernes* 337
Antoine Lonnet et Marie-Claude Simeone-Senelle
- 20 *Chaha (Gurage) Phonology* 373
Wolf Leslau
- 21 *Amharic Phonology* 399
Wolf Leslau

Egyptian Sub-branch

- 22 *Egyptian and Coptic Phonology* 431
Antonio Loprieno

Berber Languages

- 23 *Berber Phonology* 461
Maarten G. Kossmann and Harry J. Stroomer

Cushitic Languages

- 24 *Awngi Phonology* 477
Robert Hetzron

25	<i>Oromo Phonology</i>	493
	Maria-Rosa Lloret	
26	<i>Somali Phonology</i>	521
	Annarita Puglielli	
Chadic Languages		
27	<i>Hausa Phonology</i>	537
	Paul Newman	

Volume 2

2 Indo-European Languages

Ancient and Medieval

Anatolian Languages

28	<i>Hittite Phonology</i>	555
	H. Craig Melchert	

Iranian Languages

29	<i>Old Persian and Avestan Phonology</i>	569
	David Testen	
30	<i>Pahlavi Phonology</i>	601
	Dieter Weber	

Modern

Indo-Aryan Languages

31	<i>Hindi-Urdu Phonology</i>	637
	Alan S. Kaye	
32	<i>Gujarati Phonology</i>	653
	P. J. Mistry	

Iranian Languages

33	<i>Persian Phonology</i>	675
	Gernot L. Windfuhr	
34	<i>Kurdish Phonology</i>	691
	Ernest N. McCarus	
35	<i>Ossetic Phonology</i>	707
	David Testen	
36	<i>Pashto Phonology</i>	733
	Josef Elfenbein	
37	<i>Balochi Phonology</i>	761
	Josef Elfenbein	

viii Contents

	Armenian Sub-branch	
38	<i>Armenian Phonology</i>	777
	John A. C. Greppin	
3	Dravidian Languages	
39	<i>Brahui Phonology</i>	797
	Josef Elfenbein	
4	Nilo-Saharan Languages	
40	<i>Nilo-Saharan Phonology</i>	815
	M. Lionel Bender	
5	Niger-Congo Languages	
41	<i>Swahili Phonology</i>	841
	Ellen Contini-Morava	
42	<i>Sango Phonology</i>	861
	James A. Walker and William J. Samarin	
6	Altaic Languages	
	Turkic Languages	
43	<i>Turkish Phonology</i>	883
	Bernard Comrie	
44	<i>Tatar (Volga Tatar, Kazan Tatar) Phonology</i>	899
	Bernard Comrie	
45	<i>Uyghur Phonology</i>	913
	Bernard Comrie	
7	Caucasian Languages	
46	<i>Georgian Phonology</i>	929
	Howard I. Aronson	
47	<i>Chechen Phonology</i>	941
	Johanna Nichols	
48	<i>Lak Phonology</i>	973
	Gregory D. S. Anderson	
8	Unaffiliated Languages (Language Isolates)	
49	<i>Sumerian Phonology</i>	1001
	John Hayes	
50	<i>Burushaski Phonology</i>	1021
	Gregory D. S. Anderson	

List of Maps

1. Geo-Chronological Distribution of Akkadian Dialects and of Amorite	8
2. Les langues sudarabiques modernes	338
3. Nilo-Saharan Locations	818
4. The Swahili-Speaking Area	843
5. Ubanguian Languages	862

List of Tables

1-1. Synopsis of Consonants and Glides (semivowels) for Amorite, Old Akkadian, and Later Akkadian Dialects (OB+)	16
1-2. Correspondences among Laryngeals and Pharyngeals	18
1-3. Graphemic Sets and Modern Phonemic Notations for Sibilants with Divergent Realizations	20
1-4. Phonemic and Graphemic Correspondences for Sibilants with Divergent Realizations	20
1-5. Akkadian Vowels	22
1-6. Amorite Vowels	23
4-1. Phoenician Consonants	59
4-2. Phoenician Vocalic System	60
6-1. Tiberian Hebrew Vowels	91
7-1. Jewish Palestinian Aramaic Consonants	105
8-1. Old Aramaic Consonants	119
9-1. Syriac Consonants	132
10-1. Mandaic Segments	142
11-1. Old South Arabian Consonants	163
11-2. Semitic Consonant Correspondences	166
12-1. Ge'ez Consonant Correspondences	173
12-2. Ge'ez Consonant Articulation	174
12-3. Ge'ez Vowels	177

x List of Tables

13-1.	Comparison of the Major Semitic Languages	188
13-2.	Modern Standard Arabic Consonants	192
14-1.	Primary Moroccan Colloquial Arabic Consonants	208
15-1.	The Cypriot Arabic Segmental Paradigm	222
15-2.	Cypriot Arabic Dental Stop Contrasts	233
16-1.	Maltese Consonants	249
16-2.	Maltese Short Vowels	264
16-3.	Maltese Long Vowels	268
16-4.	Maltese Diphthongs	270
17-1.	Israeli Hebrew Consonants	287
18-1.	Amadiya Aramaic Consonants	315
19-1.	Correspondances consonantiques en sémitique méridional . . .	346
19-2.	Système consonantique du sudarabique moderne commun . . .	349
20-1.	Chaha Vowels	395
21-1.	Amharic Consonants	400
21-2.	Amharic Vowels	419
22-1.	Early Egyptian Consonants	437
22-2.	Early Egyptian Vowels	440
22-3.	Early Egyptian Syllabic Structures	441
22-4.	Later Egyptian Vowels	444
22-5.	Later Egyptian Syllabic Structures	445
22-6.	Coptic Consonants	447
22-7.	Sahidic Coptic Vowels	452
22-8.	Sahidic Coptic Syllabic Structures	455
23-1.	Tashelhit Consonant Inventory	467
24-1.	Awngi Consonants	478
24-2.	Awngi Vowels	482
25-1.	Oromo Vowels	494
25-2.	Oromo Consonants	495
25-3.	Oromo Transcriptions	495
26-1.	Somali Consonants	522
26-2.	Somali Vowels	524
26-3.	Somali Sandhi	532
27-1.	Hausa Consonants	538
27-2.	Hausa Vowels	541
28-1.	Hittite Consonants	559
29-1.	Iranian Languages	570
29-2.	Common Iranian Phonemes	574

29-3. The Old Persian Writing System	578
29-4. Old Persian Phonemes	580
29-5. The Avestan Writing System	587
30-1. Pahlavi Phonemes	611
31-1. Hindi-Urdu Consonants	642
31-2. Hindi-Urdu Vowels	642
31-3. Comparative Indo-Aryan Phonology	647
32-1. Gujarati Consonants	656
32-2. Gujarati Vowels	657
33-1. Persian Consonants	677
33-2. Persian Vowels	678
33-3. Persian Consonant Classification	678
33-4. Persian Glides and Vowels	678
34-1. Consonant Phonemes	692
34-2. Vowel Phonemes	696
35-1. Consonantal Phonemes	710
35-2. Semivocalic Phonemes	711
35-3. Vocalic Phonemes	711
35-4. Transitive Past Tense	714
35-5. Intransitive Past Tense	715
35-6. Vowel Alternations in Verb Stems	726
36-1. Pashto Consonants	741
37-1. Balochi Consonants	771
38-1. Classical Armenian Consonants	783
39-1. Brahui Consonants	800
40-1. Proposed Genetic Structure of Nilo-Saharan	815
40-2. Nilo-Saharan Segments Overall	823
40-3. Common Nilo-Saharan Segments	832
40-4. Proto-Nilo-Saharan Segmental Phonemes	836
41-1. Standard Swahili Consonants	844
41-2. Standard Swahili Vowels	850
41-3. Proto-Sabaki Segments	857
42-1. Pidgin Sango Vowels	864
42-2. Sango Consonants	866
42-3. Contractions of Particles	872
42-4. Urban and Rural Features	878
43-1. Turkish Vowels	884
43-2. Turkish Consonants	885

xii List of Tables

44-1. Tatar Vowels	900
44-2. Tatar Consonants	902
45-1. Uyghur Vowels	914
45-2. Uyghur Morphophonemic Vowel Qualities	914
45-3. Uyghur Consonants	915
46-1. Georgian Stops	929
46-2. Georgian Fricatives	930
46-3. Georgian Liquids, Nasals, and Glides	930
46-4. Georgian Postvelars	931
46-5. Georgian Vowels	931
46-6. Georgian Vocalic Distinctive Features	932
46-7. Georgian Phonemic System	932
47-1. Chechen Consonants	943
47-2. Chechen Vowels	945
47-3. Chechen Umlaut	948
47-4. Chechen Umlaut: <i>i</i> -Verbs	948
47-5. Chechen Dialect Vowels	956
47-6. Chechen Dialect Umlaut Processes	959
47-7. Chechen Consonant Frequencies	966
47-8. Proto-Nakh Consonant Cognate Sets	969
48-1. Lak Vowels	974
48-2. Lak Consonants	978
49-1. Sumerian Consonants	1008
49-2. Sumerian Vowels	1011
49-3. Possible Sumerian Five-Vowel System	1012
49-4. Possible Sumerian Eight-Vowel System	1012
50-1. Burushaski Consonants	1023
50-2. Burushaski Vowels	1027

2 Indo-European Languages

This page intentionally left blank.

Hittite Phonology

H. Craig Melchert

University of North Carolina at Chapel Hill

28.1. Introduction

Hittite was the chief administrative language of the Hittite Empire. It is attested in extensive cuneiform documents dating from the middle of the sixteenth to the end of the thirteenth century B.C.E. With one notable exception (see Otten 1988), these texts are inscribed on baked clay tablets. Most of these tablets were discovered at the ancient Hittite capital of Hattusha (modern Boğazköy, now Boğazkale) in the central Anatolian plateau. However, there have also been important finds at Maşat Hüyük to the northeast and at Ugarit (Ras Shamra) on the coast of Syria. A set of Hittite texts found very recently at Ortaköy near Çorum has yet to be evaluated.

The content of the Hittite texts is quite varied. While the great majority deals with aspects of the state religion (cultic administration and practice and related myths), we also have annals and treaties, a set of legal precepts, hippological texts, rituals for personal and general use, and more (see Laroche 1971 for an overview).

The last quarter-century has seen great progress in our ability to assign a relative chronology to Hittite texts and manuscripts, and we may now distinguish three successive synchronic stages: Old, Middle, and Neo-Hittite. However, changes in the phonology in the historical period are few, and the phonological system described below may be taken as valid for all stages of the language, except where noted.

Hittite is a member of the Anatolian branch of the Indo-European family. It is thus most closely related to Luvian, Palaic, Lycian, and Lydian (attested in the second and first millennia B.C.E.) and more distantly to well-known groups such as Germanic, Italic, Slavic, and so forth. For a brief survey of comparative aspects of Hittite see Melchert 1995. In what follows I will take

AUTHOR'S NOTE: In italic cited forms, - joins transliterations of individual cuneiform signs. Acute and grave accents serve only to differentiate homophonous signs. Within square brackets or slants, . indicates syllable boundary, - indicates morpheme boundary, = indicates clitic boundary.

historical factors into account only to the extent that they help illuminate the synchronic picture of the language.

The sociolinguistic situation of the Hittite Empire was complex, and the status of the language we call Hittite as a spoken medium in the Empire at various places and times is quite indeterminate. There is some plausibility to the oft-repeated claim that the Hittite we have is a “chancellery” language for written bureaucratic use, while Luvian was the principal spoken language of the “Hittite” Empire. However, this proposal must be regarded as far from proven. In any case, I see no reason to doubt that Hittite was once a real spoken language like any other.

28.2. Writing system and orthography

The cuneiform script in which Hittite is written was borrowed from Mesopotamia, where it was employed to write (Semitic) Akkadian and earlier, of course, Sumerian. The shapes of the signs and other features of Hittite usage suggest that the system was not adopted directly from Assyrian trading colonies in central Anatolia, but passed on through a North Syrian intermediary (Gamkrelidze 1961: 414ff.). It is possible, but by no means assured, that we may identify the Hurrians as this intermediary (see below on the orthography of stops).

As employed by the Hittites, the cuneiform script is a mixed logographic-syllabic system. Some signs function as Sumerograms or Akkadograms, indicating the meaning, but not the phonetic shape, of the underlying Hittite word. One also finds such logograms with “phonetic complements,” where the grammatical endings are spelled out phonetically. Finally, many words are written entirely phonetically. Obviously, these must be the basis for our analysis of Hittite phonology.

The phonetic component of the cuneiform script is a syllabary, with signs for V, CV, VC, and CVC. The script may thus unambiguously indicate word-final consonants, but initial or final clusters or internal clusters of more than two consonants can be represented only by the use of “empty” vowels which are not to be read: e.g. *kar-(a)p-ta* = /karpta/. Obviously, for us non-native speakers, determining just which graphic vowels are to be ignored is a difficult task. There are some clues suggesting that a given written vowel is not linguistically real: (1) alternate spellings with and without a vowel or with different vowels (*ki-iš-ri-it*, *ki-iš-ša-ri-it*, *ki-iš-še-ri-it* = [kis.srid]); (2) alternate spellings with the vowel before and after a consonant (*ki-iš-šar-ta* and *ki-iš-*

šar-at = /kis.sard/); (3) “broken” spellings with (C)VC-VC (*kar-ap-ta* = /karp-ta/ above). The last argument is based on the observation that the Hittites normally spell /VCV/ sequences (C)V-CV, not VC-V (/karapta/ would be spelled *ka-ra-ap-ta*). However, none of these devices is used consistently, and scholars differ sharply on the interpretation of many individual cases.

Readers should also note that the syllabary does not indicate the voicing quality of stops in VC signs: there is thus a single sign for respectively /ab/ and /ap/, /ad/ and /at/, /ag/ and /ak/, and so on.

The Akkadian syllabary distinguishes /i/ and /e/, but is deficient in Ce and eC signs. For many combinations, a single sign serves for both vowels: thus *li* = /li/ and /le/, *ir* = /ir/ and /er/. The quality of the vowel may be disambiguated by adding a preceding or following V sign (so-called “*scriptio plena*”): *le-e* = /le:/ or *e-er* = /e:r/ vs. *li-i* = /li:/ and *i-ir* = /i:r/. However, as indicated below, the principal function of the *scriptio plena* is to indicate vowel length. This device is thus not available to distinguish short /i/ from /e/, and its use even for long vowels is optional. This widespread ambiguity in the matter of *e*- and *i*-spellings causes serious problems in determining the synchronic status of /e/ and /i/.

The syllabary has signs for /ya/ and /wa/: *ta-a-ya-* = /ta:ya-/ , *ne-e-wa-* = /ne:wa-/ . Other sequences of glide plus vowel must be spelled using the vowel signs *i* and *u/ú* or *Ci* and *Cu*: *i-e-mi* = /ye:mi/, *na-a-ú-i* = /na:wil/, *a-ni-e-* = /anye-/ , *tu-e-el* = /twe:l/. Two serious problems arise in spelling glides. First, the use of *wa* and *ya* is extended from postvocalic to postconsonantal position; i.e. /Cwa/ and /Cya/ are spelled *Cu-wa-aC* and *Ci-ya-aC*, as well as expected *Cu-aC* and *Ci-aC*: *ḫa-at-ta-lu-wa-* and *ḫa-at-tal-wa-* = /ḫatalwa-/ , *ti-(ya)-an-zi* = /tyantsi/. However, in a given case it is often impossible to prove that *Cu-wa* and *Ci-ya* do not spell real disyllabic /Cuwa/ and /Ciya/ instead. Second, in the case of an ambiguous *Ci/e* sign (see previous paragraph), it is hard to know whether to read *Ci-e* as *Ce-e* (/Ce:/) or as /Cye/ (cf. *li-e* = *le-e* /le:/ but *ḫu-la-a-li-e-* = /ḫula:lye-/). For details see Melchert 1984: 83ff.

It is self-evident that analysis of the phonology of an ancient language available only in written records depends crucially on one’s interpretation of the orthography. The preceding remarks are intended to give readers some idea of the indeterminacies we face. It will be most efficient to discuss further specific problems in connection with defining the phonemic inventory. In the following phonological sketch, I will try to distinguish between those aspects on which there is fairly general agreement and those which remain controversial.

28.3. Segmental phonology

The standard transliteration of Hittite merely mimics the conventional values assigned for Akkadian and has no independent status. In attempting to define Hittite phonemes we may reasonably use all means at our disposal: internal features of the system (particularly any alternations), etymology, the treatment of Hittite names or loanwords in foreign sources, and conversely the appearance of foreign names or loanwords in Hittite. Internal factors will be treated below. Etymological considerations are of limited value for two reasons: (1) the fact that Hittite itself is part of the basis for the reconstructed systems of Common Anatolian and Proto-Indo-European makes conclusions from the latter to some extent circular; (2) we cannot in principle control for sound changes from the reconstructed pre-stages to Hittite.

Most of the languages with which Hittite comes in contact are also written in cuneiform or systems derived from cuneiform. This fact obviously robs equivalencies between them and Hittite of any probative value. Renderings of Hittite names in Egyptian hieroglyphs do confirm the basic values of many Hittite consonants: e.g. *Muršili-* /mursi:li-/ = *mrs*; *Zalhi-* /tsalhi-/ = *trh* (see further examples in Edel 1973). For vowels we must resort to spellings in the Greek alphabet of Luvian, Lycian and Lydian words which have cognates in Hittite. The weakness here is that such evidence is not only indirect, but also nearly a millennium later than Hittite. Available testimony suggests that the phonetic values implied by the symbols chosen below are broadly correct, but readers should bear in mind the approximative nature of these values. As usual in the case of a written language, we can for the most part only recover the phonemic contrasts, not the phonetic details.

28.3.1. Consonants

See Table 28-1.

28.3.1.1. Stops

As to point of articulation, only the labiovelars call for any special comment. Evidence for unitary labiovelars comes from the lexeme /e:g^w-/ ‘drink’ (Lindeman 1965): (1) occasional spellings of the stem as *e-uk-* beside usual *e-ku-* suggest that the labial articulation is simultaneous with the velar closure; (2) the first plural present is *akweni* = [agweni], implying underlying /ag^w-weni/ and unrounding of the labiovelar before /w/ (underlying */agw-weni/ would lead to [aguweni] and then *[agumeni], as per § 28.5.3); (3) preterite third singular *ekutta* implies [e:g^wta] with final prop-vowel after two conso-

Table 28-1. Hittite Consonants

	Labial	Dental	Velar	Labiovelar	Pharyngeal
Stops					
Fortis	/p/	/t/	/k/	/kʷ/	
Lenis	/b/	/d/	/g/	/gʷ/	
Affricate		/ts/			
Fricatives					
Fortis					/ħ/
Lenis		/s/			/ʕ/
Liquids		/r/, /l/			
Nasals	/m/	/n/			

nants (from */e:gwt/ we would predict **ekut* = [e:ɡud]). For an argument that /kʷ/ and /gʷ/ had become [ku] and [ɡu] before consonant in Neo-Hittite see Oettinger 1979: 237.

The synchronic manner of articulation of the Hittite stops is perhaps the most disputed aspect of all Hittite phonology. A full discussion is not possible here, and the solution I offer below is necessarily tentative and does **not** reflect a consensus. Certain facts do seem clear. The Hittite scribes do not use the Akkadian CV signs with voiceless and voiced stops in a consistent contrastive manner, in either initial or medial position. While some words are spelled consistently with voiceless or voiced stop signs, the voicing quality does not correspond regularly to expected etymological values, nor can the distribution be accounted for by any **linguistically plausible** set of sound changes. There are also many instances of alternate spellings: *ka/ga-ne-eš-* ‘recognize’, *ta/da-a-an* ‘twice’, *pé-ta/da-an* ‘place’, *ki/gi-(i)-nu* ‘break open’, etc.

On the other hand, there can no longer be any doubt of the validity of the synchronic portion of so-called “Sturtevant’s Law” (Sturtevant 1932). There is a phonemic contrast intervocalically between single stops (spelled with voiceless or voiced stops) and geminate stops: *a-pa-* ‘that’ vs. *a-ap-pa* ‘back-(ward), again’, *na-ta/da-* ‘reed, arrow’ vs. *na-at-ta* ‘not’, *še-kán* ‘cubit’ vs. *še-ek-kán-za* ‘known’, etc. The diachronic aspect of the “Law” has now also been vindicated: most geminate intervocalic stops reflect etymological voiceless stops, while single intervocalic stops continue prehistorical voiced stops. Well-established conditioned sound changes of Common Anatolian now

account for nearly all exceptions (see Eichner 1973: 79ff., 100 n. 86; Melchert in press a with refs.). It is noteworthy that Hurrian documents in cuneiform show the same phonemic contrast between intervocalic geminate and simple stops, which equates to a contrast of voiceless and voiced stops in an alphabetic system (see Laroche 1976: 22 with refs.).

Diachronic evidence involving vowel lengthening suggests that the orthographic geminate stops are true geminates (or long), in that they close a preceding syllable (Melchert in press b). Along with other scholars (e.g., Petersen 1933: 22), I therefore conclude that Hittite has reanalyzed the inherited voiceless/voiced contrast in stops as one of fortis/lenis. A contributing factor to this reanalysis is the likelihood that Hittite has generalized voiceless (i.e. fortis) stops in word-initial position and voiced (lenis) stops in word-final position. For the former, note *ti-tti-* ‘install’ < PIE **d^heh₁-* ‘place, put’, where the unexpected intervocalic *-tt-* is most easily explained as due to a pre-Hittite reduplication **after** devoicing of the initial stop. Generalization of voicing in final stops is shown by preterite third singular *pait=aš* ‘went he’ (note single intervocalic *-t-!*) < **-t*. This change, which is already Common Anatolian, is comparable to that in Old Latin (Meillet & Vendryes 1968: 146f.). These neutralizations in initial and final position would have greatly reduced the “functional load” of voicing, and in the chief remaining position of contrast, the phonetic difference had become geminate (voiceless) vs. simple (voiced). A reanalysis does not seem farfetched under these circumstances and seems to best explain the attested orthographic practices. I here use the ordinary symbols for voiced stops for the lenis series for convenience. If the above analysis is correct, there would have been no voicing contrast in stops, and undoubtedly the lenis stops were phonetically voiced under favorable conditions.

28.3.1.2. *Affricate*

Hittite has a phoneme conventionally transliterated *z*. Since this sound is spelled with the same set of signs as a sequence of /t+s/, the most economical solution is to assume a voiceless (fortis) dental affricate /ts/. In all clear cases [ts] results from assibilation of **t* before **y* and **i*. Analogical processes make the conditioned allophone phonemic: *āšzi* /a:stsi/ ‘remains’ vs. *pargašti* /pargasti/ ‘height’ (see Joseph 1984 for details). Since there is no contrasting lenis or voiced phoneme, /ts/ is spelled indifferently as single or geminate intervocalically. The phonetic realization is undoubtedly [Vts.sV]: see § 28.5.2. For rare cases of /t+s+ts/ see § 28.5.1.

28.3.1.3. *Fricatives*

Hittite has a single sibilant. Its conventional transliteration as *š* is no evidence for a palatal or palatalized articulation, as noted by Hart 1983: 104ff. Egyptian spelling of Hittite confirms a dental: *Muršili-* /mursi:li-/ = *mrsr*, *Ḫattušili-* /ḫatusi:li-/ = *ḫtsr* (Friedrich 1960: 32; see also the examples in Edel 1973). Hittite /s/ is underlyingly simple: for synchronic gemination see §§ 28.5.2, 28.5.10. There are also some biphonemic sequences of /s+s/ resulting from prehistoric assimilations.

We turn next to the famous Hittite “laryngeal(s).” Contrary to some claims, there is clearly a synchronic phonemic contrast between orthographic *-ḫḫ-* and *-ḫ-* intervocalically: *šaḫḫan* ‘feudal service’ vs. *šāḫan* ‘clogged’. Transcriptional evidence suggests that the contrast is one of voicing: the city name *Ḫattuša-* appears in Ugaritic as *Ḫt*, but *tarḫuntašša-* ‘of the Storm-god’ as *trḡ(n)ds* (see Starke 1990: 144 and compare the Egyptian examples in Edel 1973). In general, however, the geminate and simple variants are determined by the same prehistoric rules which govern voiceless and voiced stops (references in § 28.3.1.1). I therefore assume in parallelism to the stops a voiceless fortis fricative which is generalized word-initially and realized as a geminate intervocalically, and a lenis fricative which is probably indifferently voiceless or voiced intervocalically. I emphasize, however, that a voiceless/voiced contrast is quite possible.

I follow Keiler (1970: 89ff.) in assuming that PIE **h*₂, the principal source of Hittite *-ḫḫ-* and *-ḫ-*, was a pharyngeal fricative. In the absence of any compelling counterevidence, I assume the same point of articulation for Hittite, hence contrasting fortis and lenis pharyngeal fricatives for which I again use the voiceless/voiced symbols merely for convenience: *šaḫḫan* = /saḫan/ vs. *šāḫan* = /sa:ḫan/. However, the corresponding phonemes in Lycian, x and g, apparently represent back velar or uvular voiceless stop [k/q] and voiced fricative [ɣ/ʁ] respectively. A similar shift in articulation in Hittite obviously cannot be excluded.

28.3.1.4. *Liquids and nasals*

All sonorants are simple. I interpret all occurring geminates as biphonemic sequences, whether these result from prehistoric or synchronic assimilation (for a case of the latter see § 28.5.10). One should note the absence of word-initial /r/. I list /r/ as dental by convention. There is a single case of apparent confusion of /r/ with /h/ (*waḫnu-* ‘turn’ for *warnu-* ‘burn’). This raises the possibility that Hittite /r/ is uvular (cf. Kronasser 1966: 70).

28.3.2. *Vowels*

/i/, /i:/	/u/, /u:/
/e/(?)	
/e/, /e:/	/a/, /a:/

It is clear that Hittite has at least four distinct vowel phonemes: /i/, /e/, /a/, and /u/. The status of the first three has never been in doubt: compare *ḫarkun* ‘I perished’, *ḫargan* ‘destruction’ (acc. sg.), *ḫarkin* ‘bright’ (anim. acc. sg.), respectively /hargun/, /hargan/, /hargin/. For the contrast of /i/ and /e/ at all periods see Melchert 1984: 78–156 and compare *ēd* ‘eat!’ vs. *īt* ‘go!’, /e:d/ vs. /i:d/. The phonetic realization of /i/, /u/, and /a/ is not likely to be far removed from that implied by those symbols. For various reasons, our control of the sound represented by /e/ is far less certain, and [e], [ɛ], or [æ], or some combination of these, is quite possible.

There is no merit to the claim that Hittite has an /o:/ phoneme distinct from /u:/. The signs *u* and *ú* are both used to spell /u(:)/: see Kronasser 1966: 19 and Melchert in press b contra Hart 1983: 124ff. et al.

After much controversy, there is now widespread agreement that the principal function of “*scriptio plena*” (reinforcing a CV or VC sign with corresponding V sign) is to mark phonemic vowel length in Hittite: *a-ap-pa* = /a:pa/ ‘back(ward), again’ but *at-ta-* = /ata-/ ‘father’ (see Kimball 1983 passim, Oettinger 1979: 233, Eichner 1980: 154 n. 77, and Melchert in press b). For reasons discussed below in §§ 28.4, 28.5.5–6, I take most phonetic long vowels to be underlyingly short. The phonemic long vowels listed above are thus all relatively rare.

It is uncertain whether we should assume a close /e/ distinct from ordinary /e/. I made this claim in Melchert 1984: 141ff., based on several conditioned prehistoric changes and an apparent incipient merger of /e/, but not /e/, with /i/ in late Neo-Hittite. The evidence remains slim, though not yet contradicted.

Hittite has a bilabial glide /w/ and palatal glide /y/. The synchronic rules involving /w/ cited in § 28.5.3–4 suggest that the glides should be regarded as underlying phonemes, not variants of /u/ and /i/.

28.4. *Accent*

Our knowledge of the Hittite accent is predictably meager. Prehistoric changes involving the voicing or “lenition” of stops (Eichner 1973, Melchert in press a, with refs.) and lengthening of vowels (Kimball 1983, Eichner 1986: 13, Melchert in press b) would seem to suggest a stress or “ictus”

accent. Since some of the rules lengthening vowels remain as living synchronic rules, we may reasonably attribute stress to the attested language. That accented syllables are also higher in pitch than unaccented syllables is likely.

28.5. Synchronic rules

The concept of synchronic phonological rules is relatively new to the study of ancient written languages such as Hittite. The following summary should be regarded as a tentative first step.

28.5.1. *s-Epenthesis*

An underlying sequence of dental stop plus dental stop or affricate is broken up by an inserted [s]: cf. present 1st pers. sg. /'e:d.mi/ > ['e:d.mi] vs. imperative 2nd pers. pl. /'e:d.ten/ > ['e:ts.ten] > ['e:ts.sten] spelled *ēz(zaš)ten* (see § 28.5.2 for the last step) and present 3rd pers. sg. /'e:d.tsi/ > ['e:ts.tsi] spelled *ēzza(z)zi*. Remarkably, this rule is a survival from Proto-Indo-European (see Mayrhofer 1986: 110ff.), which is also attested in Luvian. It is possible that the rule is finally lost by Neo-Hittite due to paradigm leveling, but compelling evidence for this is lacking.

28.5.2. *s-Gemination*

Underlying /s/ is geminated next to another consonant at syllable boundary. As suggested by Bernabé Pajares (1973: 445ff.), this is most plausibly interpreted as spreading of the /s/ across the syllable boundary. The clearest examples are next to sonorants: cf. /sas-/ 'sleep' in pres. 3rd pl. *šašanzi* [sa.'santsi] vs. *šaš(ša)nu-* [sas.'snu-] 'cause to sleep'; conversely /kars-/ 'cut' in *kar(aš)šanzi* [kars.'santsi]. For an example next to an obstruent note *ekušši* 'you drink' = ['e:g^w.si] < /'e:g^w.si/. Spellings like *e-ez-za-aš-te-en* beside *e-ez-(za)-te-en* suggest that the same rule applies to the epenthetic [s] of the rule in § 28.5.1.

28.5.3. *w-Dissimilation*

/w/ is dissimilated to [m] before or after /u/: anim. acc. pl. /pargaw-us/ > [pargamus], pres. 1st pl. /arnu-weni/ > [arnumeni] (Kammenhuber 1969: 137, Melchert 1984: 22ff.).

28.5.4. *w-Insertion*

Contrary to earlier claims (Melchert 1984: 29, Kimball 1983: 481ff., et al.), I now believe that spellings like *šu-ú-u* 'full' and *mi-(i)-ú-uš* 'gentle' represent

[*'su.wu*] and [*'mi.wus*], with a sequence [*wu*] formed historically after the dissimilation rule in § 28.5.3 (due to loss of intervocalic **y* and “laryngeals”). If this is true, then we must assume synchronic underlying forms with a hiatus, which is filled next to /*u*/ with [*w*]: /*'su.u*/ > [*'su.wu*], /*'su.a.wus*/ > [*'su.a.mus*] > [*'su.wa.mus*], /*'mi.us*/ > [*'mi.wus*]. I cannot pursue the details here, but a synchronic morphological analysis of roots /*'su-*/ and /*'mi-*/ plus adjective-forming suffix /*-u-*, /*-aw-*/ plus case endings seems quite justifiable on independent structural grounds. Note that the preceding dissimilation rule in § 28.5.3 must also apply synchronically before the hiatus-filling rule.

28.5.5. *Accent shift with clitic*

There is limited evidence for a rightward shift of the accent when a clitic is added: cf. usual *kuitman* ‘until’ = [*'kwid.man*] < /*'kwid.man*/ vs. *kuitmān=aš* ‘until he’ = [*kwid.'ma.nas*] < /*kwid.'ma.nas*/. The shift in accent is proven by the lengthening of the /*a*/ in accented open syllable in *kuitmān=aš* vs. *kuitman*, where the /*a*/ is unaccented (see § 28.5.7). I concede that we do not find orthographic evidence for this rule as frequently as we would expect, but I believe there are enough examples to show that it is real. This accent shift is, of course, reminiscent of that in Latin with enclitics such as *-que*: *uirum* = [*'wi.rum*] vs. *uirum=que* = [*wi.'rumkwe*]: see Meillet & Vendryes 1968: 128.

28.5.6. *Accent shift in yes/no questions*

As per Watkins 1982, at least one means of marking yes/no questions appears to consist of lengthening the final syllable of the sentence (perhaps also with raising of the pitch): cf. *ḫar-te-ni-i* ‘Do you (pl.) keep?’ vs. usual *ḫar-te-e-ni* ‘you (pl.) keep’. Note that in the interrogative form not only is the final /*i*/ lengthened, but the normally long [e:] of the preceding syllable is shortened. This fact argues that both vowels are underlyingly short, being lengthened under the accent in an open syllable. In yes/no questions the normal word accent of the final word is overridden by a rule which accents the sentence-final syllable: regular /*ḫar.'te.ni*/ > [*ḫar.'te:ni*] vs. interrogative /*ḫar.te.'ni*/ > [*ḫar.te.'ni:*].

28.5.7. *Vowel lengthening in accented open syllables*

For this rule as a prehistoric change in Hittite see Kimball 1983 *passim* and Eichner 1986: 13. For its status as a synchronic rule see §§ 28.5.5 and 28.5.6.

28.5.8. Vowel length in closed syllables

Here the relationship between synchronic and diachronic processes is complex. Historically, unaccented inherited long vowels were shortened (Eichner 1986: 13), but new cases of unaccented long [e:] were created. Accented short */o/ and */e/ were lengthened in closed syllables, but */a/, */i/, and */u/ were not (Melchert in press b contra Kimball 1983).

The result is that for /a/ we have synchronic alternations between accented [a:] and unaccented [a]: *ār-ti* ‘you arrive’ = [ʼa:r.ti] vs. participle *ar-ānza* ‘arrived’ = [a.ʼra:nts]. We cannot have a synchronic rule lengthening short [a] in an accented closed syllable, because we find short [a] in this environment: *arḥa* ‘away’ = [ʼar.ha] (note that if the second syllable were accented, the word would have to be **arḥā* [ar.ʼha:] by § 28.5.7). However, I know of no sure cases of unaccented long [a:]. We may therefore suppose underlying long /a:/ in morphemes such as /a:r-/ ‘arrive’ and assume that /a:/ is shortened when unaccented.

For /e/ there are also a very few examples of synchronic alternation of accented [e:] and unaccented [e]: preterite 3rd pl. ending [-e:r] vs. [-er]. Recall that there are also cases of unaccented long [e:]: animate nom. pl. ending [-e:s]. We therefore cannot have a shortening of unaccented /e:/ as we assumed for /a:/. For the few morphemes with alternating [e:/e] we can only assume underlying short /e/ and lengthening under the accent in closed syllable (the historical rule in this case). For the large number of accented long [e:] in closed syllables which do not alternate with short [e], one can either assume underlying short /e/ and application of the above rule or simply underlying /e:/.

28.5.9. Anaptyxis in initial clusters(?)

This rule is far from assured, but alternate spellings such as *ga-ra-a-pí* and *gi-ri-pa-an-zi* ‘devour’ suggest that the first vowel of this stem is “empty”, with underlying /kr-/ (see § 28.2). When we then also find occasionally *ga-a-ra-pí*, I see no alternative but to suppose that at least for some speakers there is a synchronic anaptyxis in initial clusters of velar plus liquid: /ʼkra.bi/ > [ʼka.ra.bi] > [ka:.ra.bi] (by § 28.5.7). That the anaptyctic vowel draws the accent is not surprising, since this also takes place in prehistoric anaptyxis in similar clusters: cf. **trí-* ‘three’ > *tēri-* (synchronic underlying /ʼte.ri-/ > [ʼte:.ri-] again by § 28.5.7). It is likely that the synchronic anaptyxis seen in *gārapí* also occurs in initial clusters of /sr-/ and /sl-/: cf. *šālig-* beside *šalig-* ‘accost’ and *šārap-* beside *šara/ip-* ‘sip’.

28.5.10. Assimilation of /n/ before clitic boundary

In Old Hittite, word-final /-n/ assimilates to the initial /s-/ or /m-/ of a following enclitic: *na-aš-ša-an* = [nas.san] for /n=an=san/ (conjunction plus ‘him’ plus particle); *ma-a-am-ma-an* = [‘ma:m=man] < /‘ma:n=man/ (‘if’ plus irrealis marker). While a thorough study has yet to be made, we seem to find only *na-an-ša-an* and *ma-a-an-ma-an* in Neo-Hittite. It cannot yet be determined whether this reflects loss of the rule or merely “morphophonemic” spellings.

References

- Bernabé Pajares, Alberto. 1973. “Geminación de s y sonantes en Hetita.” *Revista española de lingüística* 3: 415–56.
- Edel, Elmar. 1973. “Hethitische Personennamen in hieroglyphischer Umschrift.” In Erich Neu and Christel Rüster, ed., *Festschrift Heinrich Otten*, pp. 59–70. Wiesbaden: Harrassowitz.
- Eichner, Heiner. 1973. “Die Etymologie von heth. *mēhur*.” *Münchener Studien zur Sprachwissenschaft* 31: 53–107.
- . 1980. “Phonetik und Lautgesetze des Hethitischen: ein Weg zur ihrer Entschlüsselung.” In Manfred Mayrhofer et al., ed., *Lautgeschichte und Etymologie*, pp. 120–65. Wiesbaden: Reichert.
- . 1986. “Die Akzentuation des Lydischen.” *Die Sprache* 32: 3–21.
- Friedrich, Johannes. 1960. *Hethitisches Elementarbuch I*, 2nd ed. Heidelberg: Winter.
- Gamkrelidze, Thomas. 1961. “The Akkado-Hittite Syllabary and the Problem of the Origin of the Hittite Script.” *Archív Orientální* 29: 406–18.
- Hart, Gillian. 1983. “Problems of Writing and Phonology in Cuneiform Hittite.” *Transactions of the Philological Society* 100–54.
- Joseph, Bryan. 1984. “A Note on Assibilation in Hittite.” *Die Sprache* 30: 1–15.
- Kammenhuber, Annalies. 1969. “Hethitisch, Palaisch, Luwisch und Hieroglyphenluwisch.” In *Altkleinasiatische Sprachen*, ed. B. Spuler (Handbuch der Orientalistik division 1 vol. 2 part 1–2 fascicle 2), pp. 119–357. Leiden: Brill.
- Keiler, Allan R. 1970. *A Phonological Study of the Indo-European Laryngeals*. The Hague: Mouton (Janua Linguarum Series Practica 76).
- Kimball, Sara. 1983. *Hittite Plene Writing*. Ph.D. dissertation, University of Pennsylvania, Philadelphia.

- Kronasser, Heinz. 1966. *Etymologie der hethitischen Sprache*. Wiesbaden: Harrassowitz.
- Laroche, Emmanuel. 1971. *Catalogue des textes hittites*. Paris: Klincksieck.
- . 1976. "Glossaire de la langue hourrite." *Revue hittite et asianique* 34.
- Lindeman, Fredrik. 1965. "Note phonologique sur hittite *eku*- 'boire'." *Revue hittite et asianique* 23: 29–32.
- Mayrhofer, Manfred. 1986. *Indogermanische Grammatik*, vol. 1, part 2, *Segmentale Phonologie des Indogermanischen*. Wiesbaden: Harrassowitz.
- Meillet, Antoine, and Joseph Vendryes. 1968. *Traité de grammaire comparée des langues classiques*, 4th ed. Paris: Champion.
- Melchert, H. Craig. 1984. *Studies in Hittite Historical Phonology*. Göttingen: Vandenhoeck & Ruprecht (*Zeitschrift für vergleichende Sprachforschung* Ergänzungsheft 32).
- . 1995. "Indo-European Languages of Anatolia." In *Civilizations of the Ancient Near East*, ed. Jack Sasson et al., pp. 2151–59. New York: Scribners.
- . in press a. "The Middle Voice in Lycian." *Historische Sprachforschung*.
- . in press b. "Hittite Vocalism." In *Studi di grammatica ittita*, ed. Onofrio Carruba and Siliva Luraghi. Pavia: IUCULANO.
- Oettinger, Norbert. 1979. *Die Stammbildung des hethitischen Verbums*. Nuremberg: Hans Carl.
- Otten, Heinrich. 1988. *Die Bronzetafel aus Boğazköy*. Wiesbaden: Harrassowitz (Studien zu den Boğazköy-Texten Beiheft 1).
- Petersen, Walter. 1933. "Hittite and Tocharian." *Language* 9: 12–34.
- Starke, Frank. 1990. *Untersuchung zur Stammbildung des keilschriftluwischen Nomens*. Wiesbaden: Harrassowitz (Studien zu den Boğazköy-Texten 31).
- Sturtevant, Edgar. 1932. "The Development of Stops in Hittite." *Journal of the American Oriental Society* 52: 1–12.
- Watkins, Calvert. 1982. "Florilegium morphologicum Anatolicum of Hittite and Indo-European Studies." Paper presented at the First East Coast Indo-European Conference, Yale University, New Haven, Conn.

This page intentionally left blank.

Old Persian and Avestan Phonology

David Testen

University of Chicago

29.1. The Iranian languages

The Iranian¹ language group developed out of the Indo-Iranian branch of the Indo-European family, probably at some point late in the third millennium B.C. Since that time the Iranian ancestral language has given way to a large number of languages with widely differing structures. Among the modern Iranian languages, Modern Persian (*Fārsī*) in Iran (as well as its very close relatives *Tājikī* in Tajikistan and *Darī* in Afghanistan) and Pashto in Afghanistan are official state languages. Other significant modern Iranian languages (in terms of the number of speakers) include Kurdish and Balochi.

The traditional chronological taxonomy of the Iranian languages makes a three-way distinction between “Old Iranian” (ending with the fall of the Achaemenid Empire in the fourth century B.C.), “Middle Iranian” (from ca. 300 B.C. to ca. 900 A.D.), and “New Iranian” (the modern languages). Iranians first appear in historical sources in the ninth century B.C., when western Iranian tribes were living on the eastern frontier of the Assyrian and Babylonian empires of Mesopotamia. When the Persian Cyrus II the Great, a client of the kingdom of the Medes, broke with his suzerain in 550 B.C., the southwestern region (*Parsā*) of what is now Iran became the center for the first Persian empire. The Achaemenid dynasty united western Iran and expanded its power over the Middle East and a large area of Eurasia until its destruction by Alexander the Great in his campaigns of 331–330 B.C.; under the succeeding empires of the Arsacids in the northwest (171 B.C.–A.D. 224) and the Sasanids in the southwest (A.D. 224–637), western Iran maintained a major position on the historical stage. Despite its conquest by the Arabs in the mid-seventh century, it retained a crucial position in the political and cultural history of the Islamic world.

1. The modern name of the nation “Iran” comes from Middle Persian *Ērān-šahr*, in which *ērān* is ultimately to be traced back to **aryānām* ‘of the Aryans’. The term **arya-* and its derivatives is well attested throughout the Indo-Iranian languages and appears to have been an early self-designation of the Indo-Iranians.

Throughout much of history, major portions of the eastern Iranian area were subordinated to the empires of western Iran. During the first millennium A.D. the commercial and cultural centers of Transoxiana and the Tarim Basin left important literary legacies in the Soghdian and Khotanese (“Saka”) languages; royal inscriptions from the principality of Bactria have also survived. The steppe lands of the north, stretching from eastern Europe to Central Asia, were long dominated by Iranian nomads such as the ancient Scythians, Sarmatians, and Massagetae and the medieval Alans, until these gave way before the emerging power of the Turkic nomads.

Table 29-1 lists the most important of the Iranian languages. The bracketed names indicate extinct languages which are unattested or only indirectly documented.

Table 29-1. Iranian Languages

	Western Languages	Eastern Languages
Old Iranian	Old Persian [Median]	(Avestan) ^a [Scytho-Sarmatian group]
Middle Iranian	Pahlavi (Middle Persian) Parthian	Soghdian Khotanese/Tumšūq Bactrian [Khwarezmian] [Alanic]
New Iranian	Modern Persian/Tajik/Dari Tat Luri Kurdish Balochi Talysh Dimli Baškardi Caspian languages (Gilaki, Mazandarani, etc.)	Pashto Pamir languages (Yazghulami, Wakhi, Munji, Yidgha, Iškašmi, Sangličī, Zebaki, the Shughni group) Ossetic Yaghnobi

a. The area in which the Avestan language was originally spoken is unknown, but it is believed to have been somewhere in eastern Iran.

29.2. From Indo-European to Iranian

29.2.1. The phonological development of Indo-Iranian

Iranian shares with the Indic languages a number of phonological developments, among which the following are the most significant:

(1) The loss of the distinction between the I.E. vowels **e*, **o*, and **a* (which merge as In.Ir. **a*) and between **ē*, **ō*, and **ā* (which merge as **ā*).

I.E.	Sanskrit		Avestan	
<i>*ekuo-</i> >	<i>aśva-</i>	‘horse’	<i>aspa-</i>	id.
<i>*mātē(r)</i> >	<i>mātā</i>	‘mother’ (nom.sg.)	<i>mātā</i>	id.

(2) Brugmann’s Law, whereby I.E. short **o* developed into In.Ir. **ā* (rather than **a*) when it was located in an open syllable followed by a sonorant (cf. Sanskrit *cakāra* ‘has done’ (3sg.perf.) < **k^we-k^wor-e*).²

(3) I.E. **ə* (i.e., the syllabified manifestation of the I.E. “laryngeal” phonemes) ultimately became **i* (or **a* when followed by **y*) in the In.Ir. languages—cf. Sanskrit *pitar-* ‘father’ < I.E. **pəter-* (= **pH_ater-*).³

(4) The development of the syllabic nasals **m̥* and **n̥* into **a*, and of **m̄* and **n̄* (< I.E. **mH*, **nH*) into **ā*.

<i>*k̥mtom</i> >	<i>śatam</i>	‘hundred’	<i>satəm</i>	id.
<i>*s̥nt-</i> >	<i>sat-ya-</i>	‘true’	<i>haiθ-ia-</i>	id.
<i>*ǵn̄-to-</i>	<i>jā-ta-</i>	‘born’	<i>zā-ta-</i>	id.

(5) The shifting of I.E. **s* to **š* (and of **z*, the voiced allophone of **s*, to **ž*) when preceded by a “guttural” (i.e., an I.E. palatal, velar, or labiovelar stop), **r/r̥*, **y/i*, or **v/u*.⁴

<i>*uidus-os</i>	<i>viduṣ-aḥ</i>	‘knowing’ gen.sg.	<i>vīduš-ō</i>	id.
<i>*mizdho-</i>	<i>mīdha-</i>	‘reward’	<i>mīžda-</i>	id.

Indo-Iranian also acquired **š* and **ž* from I.E. palatal stops in clusters with dentals (e.g., Skt. *aṣṭā* ‘eight’, L.Av. *ašta*, Modern Persian *hašt* < **oktōu*).

2. This is Kleinbans’ reformulation of Brugmann’s original law (cf. Collinge 1985: 13–21).

3. The In.Ir. vocalic reflexes of the I.E. **ə* remained distinct from the inherited **i* (and **a*), however. In Iranian the reflex of **ə* was lost in medial position—cf. the compound-form **-fya-* (in the Avestan compound infinitive *ā-fīiei-δiiāi*) ‘to support’ vs. non-compounded **paya-* ‘watch’ (pres. stem) < **pə-ie-*. The I.E. “laryngeal” set consisted of at least three distinct elements (**H_e*, **H_o*, **H_a*), distinguished by their effects upon an adjacent vowel. Indo-Iranian shows no trace of these distinctions, however. See § 29.4.3.3 for problems in the Avestan reflexes of the I.E. laryngeals.

4. For the sake of simplicity we will make no distinction here between In.Ir. **š/ž* which developed from I.E. **s* after an I.E. “guttural” and those sibilants which resulted from I.E. dental obstruents in tautosyllabic clusters with “guttural” obstruents (the so-called I.E. “**p*”)—cf. Avestan *arəša-*, Sanskrit *ṛkṣa-* < In.Ir. **ṛśša-* ‘bear’ < **H_ṛkō-* (cf. Greek *arktos*).

(6) The so-called “*satəm*”-development of the I.E. palatal series into what is most likely to have originally been an affricate series (I.E. **k̥*, **ǵ*, **ǵh* > In.Ir. **č*, **ǰ*, **ǰh*),⁵ along with the merger of the I.E. velars (**k*, **g*, **gh*) and labiovelars (**kʷ*, **gʷ*, **gʷh*) as In.Ir. **k*, **g*, **gh*.

I.E.	Sanskrit		Avestan	
* <i>k̥mtom</i> >	<i>śatam</i>	‘hundred’	<i>satəm</i>	id.
*√ <i>ǵṇ-</i>	<i>jāta-</i>	‘born’	<i>zāta</i>	id.
* <i>gʷou-</i>	<i>gauḥ</i>	‘bovine’ (nom.sg.)	<i>gāuš</i>	id.

(7) The appearance of a new series of palatal obstruents (**č*, **ǰ*, **ǰh*), which arose in Indo-Iranian through the palatalization of the velar series (< the I.E. velar and labiovelar series) by a following I.E. **e* or **y/i*. This series became phonemicized as a result of the subsequent shift of **ě* > **ǣ* (see (1) above).

* <i>-kʷe</i>	<i>-ca</i>	‘and’	<i>-ča</i>	id.
* <i>gʷenH_a-</i>	<i>janiḥ</i>	‘woman’ (nom.sg.).	<i>jainiš</i>	id.

(8) By Bartholomae’s Law, clusters consisting of a voiced aspirate obstruent (**bh*, **dh*, **ǰh*, **gh*) followed by a voiceless stop or by the sibilant **s/š* became voiced throughout, the final element of the cluster becoming aspirated (e.g., Sanskrit *buddha*- ‘awakened’ ← **budh-* + *ta-*).⁶

29.2.2. The phonological developments of early Iranian

The general phonological characteristics distinguishing early Iranian from early Indic were the following:

(9) The merger of the voiced and voiced aspirate series (Common Iranian **b*, **d*, **ǰ*, **ǰ*, **g* vs. In.Ir. **b*, **bh*, **d*, **dh*, **ǰ*, **ǰh*, **ǰ*, **ǰh*, **g*, **gh*)—cf. Avestan *bara-* ‘carry’, Modern Persian *bar-* vs. Sanskrit *bhara-*, Avestan *garəβnā-* ‘seize’ (pres. stem) (< **grβnā-*) vs. Sanskrit *grbhñā-* ‘grasp’.⁷

(10) The development of the voiceless stops **p*, **t*, **k* into the continuants **f*, **θ*, **x*, resp., when followed directly by a consonant (cf. Avestan *friia-*

5. These phonemes are still affricates in the Nuristani languages in many positions—Kati *duc* ‘ten’ (where *c* = /ts/) = Sanskrit *daśa*, Avestan *dasa* (< In.Ir. **daća* < I.E. **deḱm*). In Sanskrit **č* became a palatal sibilant *ś*, **ǰ* became a palatal obstruent *j*, and **ǰh*, merging with **ǰh*, became *h*.

6. With the loss of the distinction between the voiced aspirate and the plain voiced series in Iranian, the effects of Bartholomae’s Law are largely erased through analogy. It is still clearly visible in Gathan Avestan, however—cf. G.Av. *aogədā* ‘said’ 3sg. (< **augdha* ← **augh-ta*) vs. L.Av. *aoxta*.

7. Due to this merger, it is impossible to determine whether early Iranian, like Indic, underwent the dissimilation of aspirates known as Grassmann’s law—e.g., Sanskrit *grbhñā-* < earlier **gr̥bhñā-*.

‘friend’ vs. Sanskrit *priya*- ‘dear’; Avestan *xšaθra*- ‘lordship’ vs. Sanskrit *kṣatra*- ‘dominion’). These allophones merged with the members of the independent phonemic series **f*, **θ*, **x*, which appeared in Iranian through the spirantization of the In.Ir. voiceless aspirates **ph*, **th*, **kh* (cf. Avestan *safa*- ‘hoof’ vs. Sanskrit *śapha*-, Avestan *xara*-, Modern Persian *xar* ‘donkey’ vs. Sanskrit *khara*-); **ph*, **th*, **kh* did not become continuants, however, when following an obstruent or sibilant (Iranian **vais-ta* ‘you know’ vs. **dadā-θa* ‘you have given’).

(10) The development of **s* (but not of **š* < **s*; see (5) above) into **h*.

In.Ir.	Avestan		Sanskrit	
* <i>sauma</i> -	<i>haoma</i> -	‘ritual drink’	<i>soma</i> -	id.
* <i>vas-yas</i> -	<i>vaṣiih</i> -	‘better’	<i>vasyas</i> -	id.

This change did not take place in clusters containing obstruents (cf. the paradigm of ‘be’: Avestan *asti*, *hanti* ‘is, are’ (3rd person), Modern Persian *ast*, *and* < In.Ir. **as-ti*, **s-anti*) or before **n*.⁸ A new independent **s* developed in Iranian from the In.Ir. cluster **śc* (< I.E. **sk*)—cf. Avestan *pərəsa*- ‘ask’ (pres. stem), Modern Persian *purs*- vs. Sanskrit *prccha*- (< I.E. **prk-ske*-).

(11) The change of dental obstruents to sibilants when followed directly by another dental obstruent.

* <i>vaid-tha</i>	<i>vōis-tā</i>	‘know’ (2sg.)	<i>vettha</i>	id.
* <i>budh-ti</i> -	(<i>duδuβi</i>)-	‘feeling	<i>buddhi</i> -	‘under-
	<i>buzdi</i> -	(of anxiety)’		standing’

(12) The retention of In.Ir. **z* and **ž*, which were lost in Indic.

* <i>mazga</i> -	<i>mazga</i> -	‘marrow, brain’	<i>majj-an</i> -	‘marrow’
* <i>miždha</i> -	<i>mižda</i> -	‘reward’	<i>mīḍha</i> -	id.

(13) The clusters **ćš* and **ǰž* were simplified to **š* and **ž*, respectively, in early Iranian, while **kš*, **gž* retained their initial velar. In Sanskrit, in contrast, the clusters **ćš*, **ǰž(h)* merged with the clusters **kš*, **gž(h)* as **kš*.⁹

* <i>ćšaitra</i> -	<i>šōiθra</i> -	‘dwelling’	<i>kṣetra</i> -	‘property’
* <i>vāǰ-ž-</i>	<i>važ-a</i> -	‘drive’	<i>vākṣ-</i>	‘transport’
(< * <i>vāǰh-s-</i>)		(aor. stem)		(aor. stem)

The sounds **š* and **ž* were also found in Iranian (in addition to the instances listed under (5) above, which were inherited from Indo-Iranian) in clusters

8. The development of word-initial **sr*- is unclear—cf. G.Av. *sraxtim* ‘corner’ acc.sg. = L.Av. *θraxtim* (= Sanskrit *srakti*-), L.Av. *θraotah*- ‘river’ (= Sanskrit *srotas*-) (Reichelt 1978: 52).

9. **Kš* is the Sanskrit reflex of both voiced and voiceless clusters of this sort. Other dialects of Indic, however, appear to have maintained the voicing opposition in such clusters—cf. Sanskrit *kṣara-ti* ‘flows’ vs. Prakrit *ǰharāi* ‘drips’, Bangla *ǰharā* ‘ooze out’, Assamese *zariba* ‘ooze out’. The In.Ir. stem was **gžhar*- (or **gžhal*-)—cf. Avestan *γžar*-.

with **n*—e.g., (**frašna-* ‘question’ < **praćna-*)—and with labial stops—e.g., **āf-š* ‘water’ (nom.sg.) < **āp-s*, **dibžā-* ‘betray’ (pres. stem) < **dibzha-* < **dibh-sa-* (contrast the *s* of Sanskrit *dipsa-*, the desiderative stem of *dibh-* ‘injure’).

29.2.3. *The phonemes of early Iranian*

The phonemic inventory of Common Iranian encompassed the elements shown in Table 29-2.

Table 29-2. Common Iranian Phonemes

Consonants		
<i>*p</i>	<i>*f</i>	<i>*b</i>
<i>*t</i>	<i>*θ</i>	<i>*d</i>
<i>*č</i>		<i>*j</i>
<i>*k</i>	<i>*x</i>	<i>*g</i>
<i>*ć</i>		<i>*ǰ</i>
<i>*s</i>	<i>*h</i>	
<i>*š</i>		<i>*ž</i>
Sonorants/Semivowels		
<i>*m</i>		
<i>*n</i>		
<i>*r</i> / <i>*ṛ</i>		
[*l/*!/?]		
<i>*y</i> / <i>*i</i>		
<i>*v</i> / <i>u</i>		
Vowels		
<i>*a</i>		<i>*ā</i>
<i>*i</i>		<i>*ī</i>
<i>*u</i>		<i>*ū</i>

The voiceless stops **p*, **t*, **k* had fricative allophones (**f*, **θ*, **x*) when directly followed by a consonant.

As has been noted above, In.Ir. *s changed to Iranian *h outside of the environments identified above. Early Iranian *s represents (a) the In.Ir. cluster *sč < I.E. *sĵ, (b) In.Ir. *s which was prevented from shifting to *h by virtue of being in one of the clusters described in (10) above, and (c) the sibilant which arose as a result of dental stops coming into contact (*TT > *ST).¹⁰ Iranian *z was limited, at this early stage, to positions in contact with voiced obstruents—*z-di ‘be’ (imperative) (< *s-dhi), *vřzda- ‘large’ (< *vřddha-).

Subsequently Early Iranian *č and *ǰ developed into the sibilants s and z in most of the Iranian languages. In the Southwest Iranian (Persian) area, however, these phonemes became the dental obstruents *θ and *d—cf. Avestan *fra-zānaiti* ‘learn’ (3sg.pres.) vs. O.P. *adānā* ‘knew’ (3sg.imperf.) and the Modern Persian present stem *dān-* (< In.Ir. *ǰānā-).

The liquids *r and *l merged as *r in most Iranian languages, and indeed in most of the Indo-Iranian languages in general. Sufficient traces of a distinct *l survive in both Iranian and Indic, however, to suggest the possibility that certain dialects retained a distinction between *r and *l for a considerable time (e.g., Modern Persian *lištan* ‘lick’, the Vedic cognate of which shows forms with both l and r (*leh-mi* alongside *reh-mi*)).

*I, *u, and *r were the syllabic counterparts to *y, *v, and *r respectively—cf. *iš-ta- (< *ij-ta-), past part. of *yaǰ- ‘revere’; *vid-uš-ah, gen.sg. of *vid-vāh ‘knowing’; *bř-ta- past part. of *bar- ‘carry’. Diphthongs were composed of *a or *ā followed by *i or *u (i.e., *ai, *āi, *au, *āu).

Among the vowels, a distinction of length opposed *a, *i, *u to *ā, *ī, *ū. The long *ī of Indo-Iranian was lost in Iranian, merging with *ar—Iran. *darga- ‘long’ < In.Ir. *dřgha- (Sanskrit *dřgha-*).

29.3. Old Persian

29.3.1. Introduction

“Old Persian” is the Southwest Iranian language of the royal inscriptions of the Achaemenid Empire founded by Cyrus II the Great (ruled 559–529 B.C.), who conquered the Median empire and founded a Persian state on its ruins.

10. It is often stated (on the basis of, e.g., the word for ‘fish’—Avestan *masiia-* vs. Sanskrit *matsya-*) that the In.Ir. cluster *ts resulted in Iranian *s(s), just as *tt resulted in *st. Middle Persian *mhyg* and Modern Persian *māhī*, however, in which h calls for an earlier Persian *māθ(i)ya-, suggest that this cluster merged with Iranian *č. A merger of *ts with *č makes sense if *č corresponded phonetically to an affricate ts at that time. Compare, from the Nuristani group, the Kati cognate *maci*, in which In.Ir. *-ts- has also developed as if it were In.Ir. *č (Morgenstierne 1973: 339).

Under the dynasty of the Achaemenids (O.P. *Haxāmanišiyā*), this empire at its heyday encompassed a vast area which stretched from North Africa and the Balkans to India and Transoxiana, and incorporated the ancient polities of Egypt, the Near East, Central Asia, and north India. The Achaemenid empire was destroyed in the fourth century B.C. by Alexander the Great.

The Old Persian language is known primarily through a small corpus of royal inscriptions, the best known and most extensive of which is the inscription of Bisitun (Behistun), which celebrates in Old Persian, Babylonian, and Elamite the achievements of the opening years of the reign of Darius I (ruled 521–486). In addition, information on the Old Persian language is provided by the written records of the cultures which were subject to, or in contact with, the Achaemenid state. The most important of these records were written by Elamite, Aramaic, and Babylonian scribes.

The Old Persian lexicon as it is revealed in the attested documents displays a great deal of dialectal heterogeneity, a large number of lexical items displaying characteristics of an Iranian language other than Persian. These elements are traditionally described as borrowings from the language of the Medes, who continued to play an important part in the Achaemenid empire after the overthrow of their own kingdom. Since very little is known about Median, however, the ascription of these elements to Median is a matter of plausible speculation based as much on historical as linguistic grounds.¹¹ For the most part the elements in question are characterized by the absence of phonological innovations peculiar to Persian, making it difficult to say much more than that they reflect a more conservative form of Iranian than Old Persian proper. In some cases, these conservative forms are documented alongside Old Persian cognates, as may be seen, for example, in the case of the “Median” form *vispazana*- ‘containing all peoples’ vs. the native Persian *visadana*—the “Median” form is employed in the Old Persian text, but the Persian form survives in the Elamite documentation in the form *mi-iš-šá-da-na*.¹²

Old Persian texts from the reigns of the kings following Xerxes I (486–465) begin to show the effects of developments in the direction of Middle

11. Most of the phonological developments in these non-Persian forms are common to a large number of the Iranian languages (e.g., *č > s, *čv > sp), and thus are not of any great help in identifying the source language. The sole phonological innovation among these forms which is of value in this respect is the appearance of *f* representing early Iranian *hv. See Mayrhofer (1968).

12. Gershevitch (1985), however, takes the “Median” forms to represent dialectal differentiation within Persian, rather than borrowing from outside Persian.

Persian, such as the misuse of cases and the generalization of the thematic endings at the expense of the more restricted noun-classes (e.g., acc. *būmām* ‘earth’ rather than earlier *būmim*; gen. *dārayavaušahyā*, the genitive of the name ‘Darius’, rather than the earlier *dārayavahauš*). These incorrect forms evidently reflect efforts to write in what was by then an obsolete form of the language. Presumably by this time phonological developments such as the erosion of final syllables and the simplification of vowel sequences—cf. *nīšā-dayam* ‘I established’ vs. older *niyašādayam*—were under way.

29.3.2. *The Old Persian writing system*

The interpretation of the Old Persian texts is complicated by the nature of the writing system through which they have been transmitted. The orthographic system of the Achaemenid Persian inscriptions, although clearly inspired by the cuneiform syllabary of the ancient Mesopotamian languages, is an independent development occupying a position halfway between that of a purely consonantal alphabet (like that of early Aramaic) and a syllabary based upon symbols representing consonant–vowel sequences. For a few consonants Old Persian employs discrete characters to identify the vowel following a consonant—thus distinguishing by the use of distinct characters <d^(a)>, <dⁱ>, <d^u>; <m^(a)>, <mⁱ>, <m^u>—while for other consonants it makes no graphic distinction whatsoever to reflect the following vowel, or the absence of a vowel (e.g., the single symbol <s> in principle may represent /s/ before *a*, *i*, *u*, or $\emptyset$). A number of consonants show a partial set of vowel distinctions (e.g., <t^u> vs. <t^a/tⁱ/t>). At no point does the writing system distinguish between a vowelless consonant (C) and a consonant followed by a short *a* (Ca) (Table 29-3). In addition to the phonetic characters listed in the table, Old Persian employed special ideograms to represent the notions ‘king’, ‘land’, ‘earth’, and the divine name ‘Ahuramazda’. In general it also used a special set of characters to render the numerals, rather than spelling them out phonetically.

In word-initial position, the vowel-signs <a>, <i>, <u> are used to indicate a vowel, but do not distinguish short vowels from long vowels. Outside of initial position, the symbol <a> is used to indicate a long *ā* (i.e., <a-a> = /tā/), while <i> and <u> mark the high vowels (/i/, /ī/, /u/, /ū/), regardless of length; the absence of a vowel sign indicates either short *a* or zero. The semi-vowel characters are used sporadically to mark length (e.g., <u-v^(a)> = /ū/), as well as to represent the diphthongs /ai/, /au/. The writing system thus distinguishes between high vowels and diphthongs only when these are preceded by one of the consonants which are capable of indicating the

Table 29-3. The Old Persian Writing System

	C ^(a)	C ⁱ	C ^u
p	<p>		
t	<t ^(a-i) >		<t ^u >
c	<c>		
k	<k ^(a) >	*	<k ^u >
b			
d	<d ^(a) >	<d ⁱ >	<d ^u >
j	<j ^(a) >	<j ⁱ >	*
g	<g ^(a) >	*	<g ^u >
f	<f>		
θ	<θ>		
x	<x>		
m	<m ^(a) >	<m ⁱ >	<m ^u >
n	<n ^(a-i) >		<n ^u >
r	<r ^(a-i) >		<r ^u >
l	<l>		
v	<v ^(a) >	<v ⁱ >	*
y	<y>		
ç	<ç>		
š	<š>		
z	<z>		
h	<h>		

differences lying in a following vowel: <d^(a)-i> = the diphthong /dai/ vs. <di-i> = /dī/, whereas <t^(a-i)-i> = either /tai/ or /ti/. In general <i> and <u> are used systematically whether the preceding consonant renders them redundant or not—i.e., /di/ is spelled <di-i> rather than simply <di>. The long diphthongs /āi/ and /āu/ are written using the character <a> together with the high-vowel sign: <t^(a-i)-a-i>, <t^(a-i)-a-u> = /tāi/, /tāu/. On rare occasions an extra <a> was employed in order to make explicit an intended diphthong—cf. *Cišpaiš*, the genitive singular of the name ‘Teispes’, which is

occasionally found spelled <-p-a-i-š> rather than simply <-p-i-š> in order to distinguish it from the nominative singular *Cišpiš*.

To varying degrees, the nasal consonants /n/ and /m/ in syllable-final position were systematically omitted from writing. This omission is absolute in the case of /n/ (e.g., <a-b-r^(a-i)> = /abaran/ 'they bore'). Syllable-final /m/, however, was written using the <m^(a)>-character in word-final position (including when it was followed by a clitic) and in medial position when not followed by a stop—e.g., <a-d^(a)-m^(a)-š-i-m^(a)> = /adam-šim/ 'I ... it' vs. <k^(a)-p-d^(a)-> = the Median place-name /Kampanda-/ (cf. the Babylonian spelling *ḫa-am-ba-nu* and the Elamite spelling *qa-um-pan-taš*).

The consonant *h* was omitted regularly from writing when it preceded *ū* (e.g., <u-v^(a)-s-p-> = /huvaspa-/ 'rich in horses'), and often before *m*. The reflex of the sequence **hi*, on the other hand, was normally spelled simply as <h> without a following <i> (e.g., <h-z-a-n^(a-i)-m^(a)> </hizānam/ 'tongue' (acc.sg.)). In principle it is therefore possible to distinguish between high vowels and diphthongs in the environment of a preceding *h*, since <h-i>, <h-u> will typically represent /hai/, /hau/.

The writing system of Old Persian thus left unexpressed (or at least inadequately expressed, to one degree or another) a number of important features of the language.

(1) CV vs. C readings: There was no means of indicating that a given consonant was not followed by a vowel, thus making it possible in principle to read a hypothetical graphic sequence <p-t^(a-i)-> as /pata/, /pat/, /pta/, or /pt/. The only character which escaped this ambiguity was <n^(a-i)>, which, as a result of the conventions of representing nasals, could never represent a syllable-final *n* (i.e., <n^(a-i)-t^(a-i)> = /nata/ or /nat/, but never */nta/ or */nt/. With consonantal signs in which a following vowel was inherent (such as <d^(a)>, <dⁱ>, <d^u>), the vowelless consonant was regularly expressed using the <C^(a)> character.

(2) Syllable-final *n* was never written, and syllable-final *m* could be left unexpressed under certain circumstances. In principle, therefore, any sign not followed by a vowel-sign may be read as a syllable with a nasal coda—e.g., <p-t^(a-i)>, alongside the possible readings given above, may be read as /pan-tan/, /pan-ta/, /pa-tan/, etc.

(3) The quantity of vowels is routinely visible only in the case of /a/ vs. /ā/ (<Ca> vs. <Ca-a>) outside of initial position; the use of the semivowels <y> and <v^(a)> to indicate length in the high vowels is sporadic at best—e.g., /dahyūnām/ 'lands' (gen.pl.) is attested both as <d^(a)-h-y-u-v^(a)-n^(a-i)-a-m^(a)> and as <d^(a)-h-y-u-n^(a-i)-a-m^(a)>. The vowel signs do not indicate length when

used at the beginning of a word, making it impossible to distinguish between /apa/ and /āpa/, both of which would be written as <a-p>.

(4) There is no visible distinction between a high vowel and a diphthong unless the preceding consonant has <C^(a)> vs. <Ci> or <Cu> spellings. The script will therefore make no distinction, for example, between /-tiy/, the third person singular active verbal ending, and /-taiy/, the corresponding middle-voice ending, both of which are written as <-t^(a-i)-i-y>.

The interpretation of a given Old Persian text is therefore heavily dependent on external evidence for the proper reconstruction of the lexical items and grammatical forms. This evidence is mainly derived from comparative linguistic (Indo-Iranian or Indo-European) research; in addition, evidence of considerable value may be provided by foreign (Elamite, Aramaic, etc.) sources. Reading may be facilitated, finally, by the presumed phonotactics of the language.

29.3.3. *The phonemic inventory of Old Persian*

The Old Persian phonemic inventory contained the elements shown in Table 29-4.

Table 29-4. Old Persian Phonemes

Voiceless Stops	p	t	c	k
Voiceless Continuants	f	θ		x
Voiced Stops	b	d	j	g
Nasals	m	n		
Liquids			r, l	
Sibilants			s, š, ç (ž?)	
			h	
Semivowels			y, w	
Vowels	a	ā		
	i	ī		
	u	ū		

29.3.3.1. *Obstruents*

It has been conjectured that the written forms <b, d, g> represented voiced continuants rather than stops. Such an interpretation would allow for the development of Iranian *j > O.P. <d> (i.e., by this view, /ð/) to parallel that of voiceless *č > θ.

Voiceless *p* and *k* alternated with *f* and *x*, respectively, when in contact with a following consonant (cf. *parsā* ‘examine’ (imperative) vs. *u-frašta-* ‘well punished’). Voiceless *t*, which became **θ* in early Iranian in this position, ultimately became *š* before *n* in Old Persian (*arašnīš* ‘cubits’ (acc.pl.) vs. Avestan *arəθna-* ‘elbow’), while the sequence **θr* gave rise to *ç*—cf. the alternation in *pitā* ‘father’ (nom.sg.) (stem **pitar-*) vs. gen.sg. *piça* < Iranian **piθr-ah*. In other cases an alternation has been eliminated through analogical leveling (e.g., *gāθum* ‘place’ (acc.sg.), *xraθum* ‘wisdom’ (acc.sg.) rather than **gātum*, **xratum*, showing expansion of the *θ* which arose in the sequence **-θv-* ultimately < **-tv-*).

29.3.3.2. *The sibilants and related issues*

The most conspicuous differences between the phonemic inventory of Common Iranian and that of Old Persian lie in (1) the development of Common Iranian **č* and **ǰ* into Persian *θ* and *d*, resp., rather than into the *s* and *z* found elsewhere in the Iranian languages, and (2) the appearance of a sibilant *ç* which has developed from Iranian **θr*.¹³ The phonetic nature of O.P. <*ç*> remains unclear, although its renderings in other writing systems—cf. the spellings of the royal name *Artaxšaça-* in Babylonian (*ar-ta-’ḫa-šā-is-su*) and Egyptian Aramaic (*’rthššš*)—as well as the *s* which reflects it in Modern Persian—cf. Modern Persian *se* ‘three’ < **çaya* < **θayah*—suggest that *ç* was some form of sibilant. New instances of *s* and *z* have appeared (3) in Old Persian from the Iranian sequences **čv* and **ǰv*, respectively, which developed into *sp* and *zb* in most other Iranian languages, while new instances of *š* (4) resulted from the shift of **čy* to *šy*, which also occurred in Avestan.

	Early Iranian	Old Persian		Avestan
(1)	* <i>vič-</i>	<i>viθ-am</i>	‘house’ (acc.sg.)	<i>vīs-əm</i> id.
	* <i>ǰaranya-</i>	<i>daraniyam</i>	‘gold’	<i>zarainim</i> id.
(2)	* <i>puθra-</i>	<i>puça</i>	‘son’	<i>puθra-</i> id.
	* <i>θritiya-</i>	<i>çitiyam</i>	‘third’ (adv.)	<i>θritīm</i> id.
(3)	* <i>ačva-</i>	<i>asa-bāra-</i>	‘horse-rider’ ¹⁴	<i>aspa-</i> ‘horse’
	* <i>hiǰū-/hiǰvā-</i>	<i>hazānam</i>	‘tongue’ (acc.sg.)	<i>hizuuā</i> (nom.sg.)
(4)	* <i>čyav-</i>	<i>a-šiyav-am</i>	‘I went’	<i>šauu-</i> ‘go’ ¹⁵

13. As well as from In.Ir. **čr*; judging by /niyačārayam/ ‘I rebuilt’ (DB 1.14) alongside Avestan *srāraii-*. Since elsewhere In.Ir. **č* went to O.P. *θ*, we may assume that the shift of **č* > *θ* preceded **θr* > *ç*.

14. The “Median” form *aspa-* is also found regularly in Old Persian (cf. *uv-aspa-* ‘rich in horses’ < **hu-ačva-*). Modern Persian *asb* is also a loanword from a non-Persian Iranian language.

15. Cf. the affricate which survives, e.g., in Ossetic *cæw-in* ‘I go’ < Iranian **čyav-*.

In addition, instances of *s*, *z* and of the sequences *sp*, *zb* entered Old Persian from “Median,” e.g., *vispa-zana*- ‘containing all nations’ (< **vićva-fana*-), the Persian counterpart of which, *visa-dana*-, is attested in the Elamite rendering *mi-iš-šá-da-na*.

If the form <n^(a-i)-i-j^(a)-a-y-m^(a)> (DB 2.64) has been correctly interpreted as /niž-āyam/ ‘I went forth’, then an O.P. (*)*ž* was spelled using the character employed elsewhere for /j/. On the other hand, a <š> is found in <d^u-u-š-i-y-a-r^(a-i)-m^(a)> ‘famine’ (acc.sg.), the Avestan cognate of which, *dužyāiriia*- ‘bringing a bad year’, shows *ž*.¹⁶

29.3.3.3. *h*

Despite the fact that it was not written before *u*, it is clear from developments in later Persian that Iranian **h* was not lost before **ū* in Old Persian—the spelling <u-v^(a)-j^(a)-> ‘Elam’ is thus to be read as /hūja-/, cf. modern *Xūz-istān*. Despite its absence in the written language, the <h> also still existed before *m*, as may be seen from the occasional writing of /ahmiy/ ‘(I) am’ (< Iranian **ahmi*>) as <a-h-mⁱ-y> (more frequently <a-mⁱ-y>). Iranian *xm* (< **km*>), like *-hm-*, remains unexpressed in Old Persian, assuming that <t^(a-i)-u-m^(a)-a-> /tau(h)mā/ ‘family’ is to be related to Avestan *taoxman*- ‘seed’ (cf. Modern Persian *tuxm*).¹⁷

O.P. *h* evidently had the effect of lowering a following *i*-vowel, since, aside from a few exceptional spellings, the vowel-sign <i> is not used after <h>—*hazānam* ‘tongue’ acc.sg. (<h-z-a-n^(a-i)-m^(a)>) < **hizvā*-, *xšnāsāhay* ‘know’ (2sg. subjunctive) (<h-y> from **-āhiy* < **-āhi*); contrast 3sg. *xšnāsātiy* (<t^(a-i)-i-y>).¹⁸

29.3.3.4. *Liquids*

The O.P. segment *l* is restricted to the following foreign names: *Haldita*- (personal name), *Labanāna*- ‘Lebanon’, *Dubāla*- (toponym in Babylonia), and *Izalā*- (toponym in Assyria). More frequently foreign *l* appears in Old Persian as *r*: *Bābiru*- ‘Babylon’, *Nadintabaira*- ‘Nidintu-Bel’. The character <l> is the only element in the O.P. script which is clearly modeled on a Mesopotamian cuneiform symbol, viz. *la* (Paper 1956).

16. Corresponding to Avestan *vīžibiīō* ‘households’ (dat.) (< **viž-byah* < **vić-byah*) we find O.P. /viθ-biš-ca/ (with <θ>) ‘and houses’ (instr.), which evidently results from the replacement of **ž* by the underlying θ (< Iranian **č*>) found elsewhere in the paradigm.

17. In Median, however, **-xm-* was expressed in the name of the Mede *Taxmaspāda*-.

18. Most of the exceptions are loanwords (e.g., <h-i-d^u-u-š> ‘Hinduš’). However, the reflex of **-hy-* in grammatical endings is also occasionally spelled with <i>: <a-h-(i)-y-a-y-a> ‘this’ loc.sg.fem. (Kent 1953: 68–69).

It is not clear from the O.P. spelling <a-r^(a-i)>, <-Ca-r^(a-i)> whether Iranian **r* was still a syllabic *r* in Old Persian or whether it had merged with the sequence *ar*. The fact that in Elamite renderings of Persian names the O.P. reflex of **r* is spelled with *ir*—cf. *ir-šá-ma-*, *ir-šá-um-ma-* = O.P. *Aršāma-* (personal name), *mi-ir-qa-nu-ia-ip* = O.P. *Varkāna-* ‘Hyrkanians’—implies that Old Persian still had a syllabic **r* which the script failed to distinguish from *ar*. In the present stem of the well-attested verb ‘do’ (e.g., <k^u-u-n^(a-i)-u-t^(a-i)-i-y> /kunauiy/, 3sg.pres.indic.), the Iranian stem **kṛn(a)u-* has undergone an anomalous shift of **r* > *u*, presumably under the influence of the following syllable.

29.3.4. Conditioned phonological developments within Old Persian

29.3.4.1. Consonant clusters in Old Persian

The sequence *dr* when followed by *u* was broken up by an anaptyctic *u*: *duruxtam* ‘false’ (neut.sg.) < **druxta-*, *duruvā* ‘firm’ (fem.sg.) < **druvā-*. The name for ‘Sogdia’ shows various spellings (<s-u-g^u-u-da->, <s-u-g^(a)-da->, <s-u-g^u-da->, the last without a vowel sign after <g^u>!), at least certain of which imply the presence of a vowel *u* between the *g* and the *d*. It is possible that intrusive vowels were present in other clusters but are not revealed by the script.

The semivowels **y* and **v* became vocalized after a consonant (**Cy*, **Cv* > *Ciy*, *Cuv*)—*aniya-* ‘other’ < **anya-*, *θuvām* ‘you’ (acc.sg.) < **θvām*. They thus merged with original **-Ciy-*, **-Cuv-* (cf. *tuva(m)* ‘you’ nom.sg. < **tuvam* vs. acc. *θuvām* < **θvām*).

Clusters alien to the native Persian phonological system were introduced through foreign words (e.g., *Katpatuka-* ‘Cappadocia’, *Miθra-*, *Mitra-* ‘Mithra’, *Bāxtriš* ‘Bactria’).¹⁹

29.3.4.2. Consonants in final position

The only consonants allowed in word-final position (as revealed by the script, at least) are *-š*, *-m*, and *-r*. It is not clear whether the remaining consonants were dropped or simply not expressed in the writing system.²⁰ The

19. The native Persian forms corresponding to *Miθra-*, *Bāxtriš* (i.e., *Miça-*, *Bāxçiš*) are preserved in Elamite *mi-iš-šá-*, *ba-ak-ši-iš*.

20. It is sometimes argued that the failure of vowels to lengthen before a missing consonant (see below) is evidence that the consonants were actually still present and merely unexpressed by the script. It may equally well be argued, however, that the absence of lengthening is simply due to relative chronology, assuming that the lengthening of final vowels preceded the dropping of the final consonants.

rules governing the *Auslaut* also obtained word-internally within compounds (cf. *Dāraya-va(h)uš* ‘Darius’ < **dāraya(t)* ‘holding’- **vahu-* ‘good’).

29.3.4.3. Vowels in final position

An early Iranian short *-a in word-final position was lengthened to -ā (and spelled with <a>) in Old Persian. By origin an O.P. final -ā could therefore represent (1) an original Iranian *-ā, (2) an Iranian *ā which had come to be in final position due to the loss of a final consonant, or (3) a secondarily lengthened short *-a.

(1)	<i>pitā</i>	‘father’ (nom.sg.)	< Iran. * <i>pitā</i>
	<i>hainā</i>	‘army’ (nom.sg.)	< Iran. * <i>hainā</i>
(2)	<i>draugā</i>	‘lie’ (abl.sg.)	< Iran. * <i>draugād</i>
	<i>martiyā</i>	‘men’ (nom.pl.)	< Iran. * <i>mṛtiyāh</i>
(3)	<i>martiyā</i>	‘man!’ (voc.sg.)	< Iran. * <i>mṛtiya</i>
	<i>parsā</i>	‘ask!’ (sg.imper.)	< Iran. * <i>pṛsa</i>

Consequently all instances of O.P. short -a in final position result from the loss of a final consonant: *baga* ‘god’ (nom.sg.) < **bagah*, *abara* ‘bore’ (3sg. imperf.) < **abarat*.

Final *-i and *-u are systematically spelled with <i-y> or <u-v^(a)> (e.g., <a-s-t^(a-i)-i-y> ‘is’ (3sg.pres.) < **asti*; <p-a-t^u-u-v^(a)> ‘protect!’ imper. < **pātu*). It is not clear whether this spelling represents a lengthening to -ī, -ū, or whether it is merely an orthographic convention. Note that the spelling with <y> also occurs with an -i which has come to be in final position due to the loss of a consonant (<c-i-y> < **čid* ‘what’), indicating that, if there is in fact a lengthening involved here, it differs from the lengthening of final *-a, since the latter did not affect vowels which came to be in final position through the loss of a final consonant.²¹

29.4. Avestan

29.4.1. Avestan Phonology

Avestan is the language of the Avesta, the sacred text of the Zoroastrian religion composed by the prophet Zaratuštra, who is believed to have lived around the beginning of the first millennium B.C. somewhere in the eastern Iranian lands. The text of the Avesta was transmitted orally through the Zoroastrian priestly tradition for centuries before it was committed to writ-

21. Note also that the reflex of word-final *-hi is written as <h-y>, presumably representing *-hay < *-hiy < *-hi.

ing in Persia at some point in the middle of the first millennium A.D. The oldest of the extant manuscripts of the Avesta dates back to the thirteenth century, while the most important manuscripts date no further back than the sixteenth; all the known manuscripts are believed to be derived from an original which was copied in the ninth or tenth century. There is thus an enormous span of time intervening between the composition of the Avesta and its appearance in written form, and another significant expanse of time between the first writing of the Avesta and the copying of the documents on which our knowledge of the language is based; during the majority of this time, the transmission of the text, both oral and written, was in the care of generations of priests who presumably had only limited knowledge of the language other than as a medium for ritual utterances.²²

There are two forms of the language employed in the Avesta, Gathan (or Old Avestan), which in general is more conservative, and Late Avestan. The portions of the text written in Gathan are the *Gathas* or “Hymns” (found in Yasnas 28–34, 43–51, and 53), the *Yasna haptanḡhāiti* (Yasnas 35–41), and the prayers *yaθā ahū vairiō* (Yasna 27, 13) and *ā airiōmā išiiō* (Yasna 54, 1).

Because of the complicated history of the text of the Avesta, it is difficult to ascertain exactly when and under what circumstances the developments which gave rise to the language known as “Avestan” took place, since it is often not easy to determine whether a given development took place prior to the composition of the text (and thus constitutes a development in the history of the language spoken by Zaratustra), or whether it occurred over the course of the oral transmission of the text and its appearance in written form (perhaps due to contacts between the various Iranian dialects through which the oral Avesta passed, or perhaps as interference from the native Middle Iranian language of the priests responsible for committing the text to writing), or even whether it is simply to be ascribed to the scribal transmission of the document. Since in many respects the language of the Avesta (especially Gathan) gives a general impression of great antiquity which is comparable in many respects to Vedic, the language of the prophet is frequently reconstructed with many of the phonological developments “undone”—for example, the shapes of the diphthongs *aē*, *ōi*, by the analyses of some researchers, are ascribed to developments subsequent to the composition of the Avestan, and an original Avestan **ai* reconstructed for the earliest shape

22. It has been suggested, on the basis of analyses of the Avestan writing system, that the present written form of the Avesta is based upon an earlier “Arsacid” Avesta written in a consonantal alphabet comparable to that of Middle Persian (Andreas 1904).

of the text. To a considerable degree this approach is justified by the meter of the Avesta: e.g., the diphthong of the pronoun *aēm* ‘this’ nom.masc.sg. is metrically disyllabic, indicating that at the time when the Avesta was composed the language had not yet completed the shift which ultimately eliminated the second syllable of the pronoun (**ayam* > **ayəm* > **ayim* > **aim* > *aēm*).

29.4.2. *The Avestan writing system*

The task of setting down the Avesta obliged the Sasanid priests to create a new writing system. Since the primary impetus for putting the Avesta into written form seems to have been to preserve what was held to be the correct inherited pronunciation of the sacred prayers, it was necessary to devise a script capable of rendering phonetic detail of considerable fineness. With this in mind the inventors of the Avestan alphabet requisitioned the Middle Persian (Pahlavi) alphabet available to them and made a number of crucial emendations, including the elimination of the large-scale homography which bedeviled the characters representing the Middle Persian consonants and the introduction of characters to provide a systematic reflection of the vowels. Certain of the Avestan characters were taken directly from cursive Pahlavi—<a, i, k, t, p, b, m, n, r, s, z, š, x^v> (the last being, in Middle Persian terms, a digraph <hv>)—or from the Psalter script—<γ, j, d>—while others were invented (e.g., <f> and <β> on the basis of <p>). The end result of this undertaking was an alphabetic script comprising the elements shown in Table 29-5.

29.4.2.1. *Orthographic conventions*

The character <η> reflects a nasal **n* or **m* in medial position preceding a stop (*pa<η>ča* ‘five’, *frasči<η>banəm* ‘supports’ gen.pl., *za<η>ga-* ‘ankle’); at the end of a word followed by an enclitic, final **-n* is written using either <η> or <n>, while for final **-m* in this position the character <m> is used.²³

The characters <y> and <v> are used only in word-initial position. Elsewhere the Iranian semivowels **y* and **v* are reflected by the spellings <-ii-> and <-uu->, respectively. Presumably this orthographic convention is due to a syllabification of postconsonantal **Cy*, **Cv* which gave **Ciy*, **Cuv-*, which

23. Hoffmann and Narten (1989: 67–69) suggest interpreting <η> as “der merkmallose, postuvulare, d.h. nur durch Senkung des Gaumensegels artikulierte Nasal [N], wie er etwa in der Gujurati vorkommt,” ascribing the Old Persian writing conventions concerning nasals to a similar source.

Table 29-5. The Avestan Writing System

<p>	<f>		<β>	<m>
<t>	<θ>	<d>	<δ>	<n>
<ṭ>				<ñ>
				<ṇ>
<č>		<j>		
<k>	<x>	<g>	<γ>	<ŋ>
				<ŋ̊>
				<ŋ ^v >
<s>	<š>	<ṣ>	<ś>	<h>
				<χ>
				<x ^v >
<z>	<ž>			
	<r>	<y>		<v>
<i>	<ī>		<u>	<ū>
<e>	<ē>		<o>	<ō>
	<ə>		<ā>	
	<a>		<ā̃>	
			<ā̃ ^a >	
			<ā̃>	

a. The character representing *ā̃* is formed by a combination of *ā* and *ə*. There are also traces of a short *ā̃* (i.e., in graphic terms a short *a* + *ə*) (Hoffmann and Narten 1989: 31).

merged with original **Ciy*-, **Cuv*-.²⁴ Subsequently the spellings <-ii-> and <-uu-> were employed to render the semivowels in intervocalic position as well.²⁵

The character <ā̃>, in addition to representing the nasalized reflex of Iranian **ā̃*, could evidently be used to indicate nasalization of a preceding sonorant: <nəṛāš> ‘man’ acc.pl. [to be read *nəṛəšʔ*] <**n̄ms*, <huuāmahī> [= *hūmahīʔ*] ‘we move’ <**hunmahī* (Hoffmann and Narten 1989: 72–75).

24. Note that a comparable development evidently occurred in Old Persian. The spellings <ii>, <uu> may indicate an Achaemenid Persian role in the transmission of the Avesta (Hoffmann 1975: 322–24).

25. It is also conceivable that the spelling of intervocalic <ii->, <uu-> is a faithful representation of -*VyyV*-, -*VvV*-, resulting from a shift of *-*VyV*-, *-*VvV*- to *-*VyyV*-, *-*VvV*- (i.e., with a heterosyllabic cluster resulting from gemination) (Hoffmann and Narten 1989: 39–49).

29.4.3. The phonological development of Avestan

29.4.3.1. Developments in Avestan vowels

29.4.3.1.1. VACILLATIONS IN THE REPRESENTATION OF QUANTITY

Among the characters representing the high vowels we frequently find the substitution of a long vowel for an expected short vowel and vice versa.

ī for **i* *mīždōm* (alongside *miždōm*) ‘reward’

i for **ī* *framruuiša* ‘honor’ 2sg.mid.opt. (< **fra-mruv-ī-ša*)

ū for **u* *ahūm* ‘being’ acc. (vs. nom. *aṇhuš*)

u for **ū* *buyāṭ* ‘be’ 3sg.act.opt. (vs. future stem *bū-šiia-*)

Word-final *-*im* is almost invariably spelled <-īm>: <paitīm>, <ašīm>, acc. of *paitiš* ‘lord’, *ašiš* ‘destiny’. Despite the variation in spelling, “there appears, after all, to be a certain statistical preponderance of cases, in which the ancient distinction of quantity is preserved, even with regards to *ī* and *ū*. And it does not seem probable that the original system had already been altogether abolished ...” (Morgenstierne 1973: 52).

29.4.3.1.2. CHANGES IN QUANTITY

True shifts in the quantity of certain vowels seem to have occurred, however, in the history of Avestan. The occasional appearance of *a* in the place of *ā* appears to be the result of an accentual shift caused by the addition of a clitic or a suffix (see § 29.4.3.4); for unclear reasons *ā* also appears in a number of other situations, e.g., consistently in *hāitī-* (< **hatī-*), the feminine of the participle *hant-* ‘being’, and sporadically elsewhere (Kellens 1989: 37–38).

In word-final position, oppositions of vocalic quantity are lost: in Gathan all vowels²⁶ are systematically written as long, while in Late Avestan the final vowels of monosyllables are typically written as long and the final vowels of polysyllabic words (other than -*ō*, -*ā*, -*āi*, -*āu*), as short.

	Gathan	Late Avestan
* <i>nā</i> ‘man’ nom.sg.	<i>nā</i>	<i>nā</i>
* <i>jī</i> sentence particle	<i>zī</i>	<i>zī</i>
* <i>ahura</i> ‘god’ voc.	<i>ahurā</i>	<i>ahura</i>
* <i>ahurā</i> ‘god’ inst.	<i>ahurā</i>	<i>ahura</i>

29.4.3.1.3. DEVELOPMENTS IN THE VICINITY OF NASALS

Iranian **a* regularly becomes *ə* when followed by a nasal (*azəm* ‘I’ < **azam*, *barən* ‘they bore’ < **baran*).²⁷ The sequences *-*iyaN*- and *-*uvaN*- frequently

26. Including anaptyctic *ə*—cf. *vadarə* ‘weapon’ < **vadar*.

27. The inserted nasal of -*ṇh*- < *-*h*- (see § 29.4.3.2.3) does not cause this shift. The treatment of the sequence *-*ṇh*- (< In.Ir. *-*ns*-) differs between the two dialects: for the Iranian stem **canh-* (cf. Sanskrit *śams-*) Gathan has *sāṇgha-* (with the prenasal treatment of the vowel), while Late Avestan has *saṇgha-* (the latter being comparable to the secondary -*ṇh*- < *-*h*-).

become *-īN/-ūN-* (or *-īN/-uN-*) (cf. *θritīm* ‘third’ adv. < **θritiyam*; *bun* ‘were’ 3pl. < **buvan*), presumably via **-iyəN*, **-uvəN*; in the post-consonantal sequences **-yaN-* and **-vaN-* the shift of **a* > ə may be blocked (*tbišyantəm* ‘enemies’ gen., *drəguuantəm* ‘adherent of falsehood’ acc.), or the sequence may be simplified to *-īN-*, *-ūN-* as with **-iyaN-/uvaN-* (*irišintəm* ‘injured’ gen.pl., *tafnahuntəm* ‘containing heat’ acc.sg., < **rišyantām*, **tafnahvantam*). Presumably many of the cases of *i* < **a* following *č* or *j*—such as *tačin* ‘ran’ 3pl.impf. < **tačan*, *jimaitī* ‘come’ 3sg.subj. < **jamati*—have gone through a stage of **ə* conditioned by a following nasal.

A low vowel (**ǣ*) preceding a cluster of nasal + continuant gives rise to a phonemic *ǣ*—*mǣθrəm* ‘spell’ < **manθram*—which also results from **ā* before a word-final nasal—*mām* ‘me’ acc. tonic < **mām*, *vərəziiān* ‘work’ 3pl. subjunctive < **vřjyān*. Occasionally *ǣ* is also found for *ā* in an open syllable followed by a nasal (e.g., *daδāmi* ‘I give’).

It is possible that nasalized **i* and **u* formerly existed but have been lost, assuming that the accusative plural endings of the *i*- and *u*-stem nouns (*-īš*, *-ūš*, resp., < **-ins*, **-uns*) went through a nasalized phase. New nasalized allophones *i̯* and *u̯*, caused by the following cluster *-nm-*, are in all likelihood to be seen in the G.Av. spellings <*friiānmaḥī*> ‘bless’ 1pl.pres., <*hūuānmaḥī*> ‘drive’ 1pl.pres., <*duuānman-*> ‘cloud’; similarly, reflexes of a nasal **r* (*-ərə-?*) seem to lie behind the spelling <*əra*> of forms such as *mātəraš-čā* ‘and mothers’ acc.pl.

Word-final **-anh* (< In.Ir. **-ans*) is represented by G.Av. *-əng*, L.Av. *-ə* or *-ǣ*—cf. *mašiiēng* ‘men’ acc.pl. (< **mřtiyanh*), *xvəng* ‘sun’ gen.sg. (ult. < **huanh*) vs. *mazištā amā* ‘greatest powers’ acc.pl. (< **majištanh amanh*), *hū* ‘sun’ gen.sg. (< **huuə* < **huanh*) (Hoffmann 1975: 274–87).

Iranian **ā* became *ǣ* before a cluster consisting of a nasal plus a voiceless consonant (including *-ŋh-* < **-h-*; see § 29.4.3.2.3)—*hačāntē* ‘follow’ 3pl.pres.mid. subjunctive (< **hačāntai*), *māŋhō* ‘moon’ gen. (< **māh-ah*).

29.4.3.1.4. PALATALIZATION AND VOWELS

Iranian **a* > *e* when preceded by *y-/ii-* and followed by one of the palatal elements (*-ii-*, *-j-*) or a front vowel, unless *r*, *uu*, or *hm* intervenes—G.Av. *ye-hiiā*, L.Av. *yehe/yerfhe* gen.sg. of the relative pronoun < **yahya*, L.Av. *ye(i)di* ‘when’ < **yadi*, *iθiiejō* ‘danger’ < **θyajah-* vs. *yahmi* loc.sg. of the relative pronoun.

Iranian **ā* > *e* occasionally in the environment of a palatal and a front vowel—cf. *ziieni-* ‘damage’, *ā-iiēsē* ‘takes away’ 3sg.mid. Morgenstierne (1973: 48), citing *aiienī* ‘I shall go’ < **ayāni* and *jaiðiemi* ‘I ask’ alongside

jaiḍiiāmi, observes, “Usually the accumulated effect of *y* and a following nasal + palatal vowel is needed in order to change *Ā* into *e*.”

The *ə* resulting from **a* followed by a nasal frequently becomes *i* in a palatalizing environment—cf. *yim/yəm* < **yam*, acc.sg. of the relative pronoun, *vāčim/vačəm* ‘word’ acc. < **vāč-am*.

29.4.3.1.5. LABIALIZATION AND VOWELS

Iranian **a* > *o* when flanked by a labial element and the vowel *u*—cf. *vohu* ‘good’ nom.sg.neut. vs. *varhəuš* gen.sg. < **vahu*, **vahauš*. Occasionally **ā* becomes *ō* before *u* in the following syllable: *jiiōtūm* ‘life’ acc. (vs. gen. *jiiātəuš*).

The sequence **-və-* (where **ə* < **a* preceding a nasal) may become *-u-*, especially in the word-final syllable—*haurum* ‘all’ acc.sg. < **harvam*, *drūm* ‘well’ acc.sg. < **druvam*—in a development paralleling the shift of *-yə-* > *i* before a nasal.

Sporadically **ā* appears as *ō* before *-rC*—*čōrəṭ* ‘made’ 3sg.aor., *dōrəšt* ‘had’ 3sg.aor., *θuuōrəštāra* ‘deciders’ acc.du.

29.4.3.1.6. AVESTAN *ə*

In addition to arising from **a* before a nasal (see § 29.4.3.1.3), Avestan *ə* may come from **a* before *-uuĩ-* (*səuuīšta-* ‘most powerful’, cf. Sanskrit *śaviṣṭha-*). The *ə* is also widely found representing an anaptyctic vowel inserted into consonant clusters (see § 29.4.3.2.7). In Gathan, *ə* is found for **a* followed by **-hm-* or **-nh*—*əhmā* ‘us’ acc., *sənghāmī* ‘announce’ 1sg.pres. vs. L.Av. *ahmā*, *saṅhāmi*. A lengthened *ē* appears sporadically in Gathan in the place of *ə*—*aniiēm* ‘other’ acc.sg., and has been related by Kurylowicz (see § 29.4.3.4) to the original accentuation; this long G.Av. *ə* also seems to result from **ā* in monosyllabic words beginning with a consonant cluster and ending with a nasal (*χiiēm* ‘be’ 1sg.opt. < **hyām*). The significance of such G.Av. spellings as *ēəuuā* ‘us’ du.acc. (< **āvā*) and *ēəānū* ‘after’ (< **anu*) is unclear.

29.4.3.1.7. DIPHTHONGS AND DIPHTHONGIZATION

Avestan diphthongs come from three sources:

- (1) The reflexes of the inherited Iranian diphthongs (**ai*, **au*, **āi*, **āu*).
- (2) Diphthongs arising through vowel syncope.
- (3) Diphthongs arising through epenthesis under the influence of a high vowel or semivowel in the following syllable.

(1) Each of the original Iranian sequences **ai* and **au* appears in two separate shapes in Avestan: the original **ai* appears either as *aē* or *ōi*, while the original **au* has become either *ao* or *əu*. In word-final position we find *-e* and *-ōi* for **-ai* and *-ō* for **-au*.

<i>ašōiš</i>	‘righteousness’ gen.sg.	< * <i>ṛtaiš</i>
<i>kaēibiio</i>	‘who’ dat.pl.	< * <i>kaibyah</i>
<i>tōi</i>	‘these’ nom.pl.	< * <i>tai</i>
<i>auue</i>	‘those’ nom.pl.	< * <i>avai</i>
<i>gaobīš</i>	‘cows’ inst.pl.	< * <i>gaubiš</i>
<i>gəuš</i>	‘cow’ gen.sg.	< * <i>gauš</i>
<i>mainiio</i>	‘spirit’ voc.sg.	< * <i>manyav</i>

The reason for the differentiation in the reflexes of these diphthongs remains unclear, but at least in part it appears to be related to the placement of the original accent (see § 29.4.3.4). Bartholomae (1895–1901: 172) has suggested that *aē* was originally found in word-initial position and in open syllables, while *ōi* was found in closed syllables (cf. G.Av. *vaēdā* ‘knows’ 3sg. vs. *vōistā* 2sg.), but there remain exceptions to this generalization (e.g., *vaēsma* ‘dwelling’). The *ao/əu* split of **au* is much more lopsided than that of *aē/ōi*, the *əu*-reflex being essentially restricted to the position before *š*.

The Iranian long diphthongs **āi*, **āu* are preserved as *āi*, *āu*, respectively, in Avestan (*dāiš* ‘showed’ 2sg.aor., *gāuš* ‘cow’ nom.sg.); word-final **-āu* may appear as *-ā* (e.g., *xratā* ‘will’ loc.sg.).

(2) Reflections of the diphthongs **āi*, **āu* (i.e., *aē/ōi*, *ao/əu*, *āi*, *āu*) may also appear in Avestan through the loss of a vowel, with **āi*, **āu* resulting from **āyV*, **āvV*. Most frequently this syncope is associated with a following nasal, and one may suppose a development **āyāN*, **āvāN* > **āyāN*, **āvāN* > **āyāN*, **āvāN* > **āyāN*, **āvāN* > *aēN*, *aoN*.

<i>aēm</i> (alongside <i>aiiēm</i>)	masc.nom.sg.pron.	< * <i>ayam</i>
<i>kərənaon</i>	‘do’ 3pl.subjunctive	< * <i>kṛnavan</i>
<i>haxāim</i>	‘companion’ acc.sg.	< * <i>haxāyam</i>
<i>pərəsāum</i>	‘rib’ acc.sg.	< * <i>pṛcāvam</i>

Avestan *ao* may also result from **avə* before **r* (*fraorət* < **fravṛt*). Other apparent cases of syncope are *yōišta-* ‘youngest’ < **yavišta-*,²⁸ *raēm* ‘wealth’ acc.sg. < **rayim*.

(3) New diphthongs, and even triphthongs, arise in Avestan through epenthesis. It is not clear whether the introduction of the characters <i> and <u> after an original vowel under the influence of a following syllable—e.g., *paitiš* ‘lord’ nom.sg. < **patiš*—is to be interpreted as an indication of a change in the nature of the vowel (<ai> < *<a>) or as an attempt to indicate indirectly the palatalization or rounding of the intervening consonant (<it> > = <it> < *<t>) (Morgenstierne 1973: 56–59).

28. Morgenstierne, however, reconstructs **yuvišta-*, “with substitution of *ōi* for *ui*, a diphthong which did not belong to the Av. system” (1973: 44).

The introduction of a preceding <i> occurs systematically when a medial labial or dental obstruent, *n(t)*, *ŋh*, or *r* is followed by *ĩ* or *-ii-*; final *-e* sometimes has the same effect, but the other manifestations of **ai* (*aē*, *ōi*) do not. The introduction of <i> results in alternations like the following within the paradigm.

<i>gairiṣ</i>	‘mountain’ nom.sg.	<i>garōiṣ</i>	gen.sg.
<i>gairīṣ</i>	‘mountains’ acc.pl.	<i>gara</i>	loc.sg.
<i>gairiṇam</i>	gen.pl.	<i>garaiiō</i>	nom.pl.
<i>gairibiiō</i>	dat.pl.		

When the vowel of the first syllable is *ĩ*, <i>-epenthesis results in <-ai-> rather than in **<-ii->*—*daidiiaṭ* ‘observe’ 3pl.pret. subjunctive, *daiḍiiantō* active part., nom.pl. (< redupl. **di-dyat*, **di-dyant-*) vs. *ā-diḍāiti* ‘watch’ 3sg.pres.

The appearance of <u> is regular before *r* followed by *ũ*—*auruša-* ‘white’, *aēuruṣ* ‘shining’ nom.sg., *pouruṣ* ‘much’ nom.sg. (vs. gen.sg. *paraoš*). The prefixed vowel-sign associated with *r* differs from those found with the remaining consonants in that it consistently shows epenthesis in initial as well as medial position—*irinaxti* ‘release’ 3sg.pres. (cf. Sanskrit *riṇakti*) vs. aor. stem *raēxš-*, *urūruḍiša* ‘grow’ 2sg.perf.mid. (< **ru-rud-*) vs. pres. stem *raoḍa-*.

29.4.3.2. *Developments in Avestan consonants*

29.4.3.2.1. THE IRANIAN OBSTRUENTS

The voiced stops *b*, *d*, *g* become spirantized (β ,²⁹ δ , γ) in Late Avestan, unless they are in initial position or preceded by a nasal or a sibilant,³⁰ although the presence of an intervening morpheme boundary may prevent the spirantization from taking place. In a few cases (most notably in the reduplicated stem of *dā-* ‘give, put’) an anomalous θ appears in the place of an expected δ (*daθā-* ‘give’, pres. stem, *vīθušī-* ‘knowing’ fem.).³¹

When in a cluster with a preceding fricative, the continuants *f* and θ are replaced by β , δ , respectively (e.g., *uxḍa-* ‘utterance’ < In.Ir. **uk-tha-*, *fəḍrōi* ‘father’ dat.sg. (< **ptr-ai* < **pətr-*). The cluster *-pt-* (cf. *hapta* ‘seven’), in which the original **p* has either been retained or restored, is curious.

29. Also spelled <-uu->, as if it were the reflex of the semivowel **-v-*.

30. The presence of an adjacent *r* may apparently block the shift of *d* > δ , as in *udra-* ‘otter’, *arəduš-* ‘injury’ (Hoffmann 1975: 196ff.). The existence of both *d* and δ in medial position, however, led Morgenstierne (1973: 69, 72–73) to conclude that, of the voiced continuants, *d* and δ at least were phonemically distinct in Late Avestan.

31. It has been suggested that the appearance of *daθā-* as the present stem of ‘give’ alongside *dadā-/daḍā-* is evidence for a Parthian intermediary in the transmission of the Avesta—cf. Middle Parthian <dh> < earlier **daθā-* (Hoffmann 1975: 65).

In Avestan, dental stops in word-final position³² are written with the character *t*, the significance of which is unclear (*dadāt* ‘gave’ 3sg.pret. < **dadāt*, *taṭ* ‘that’ nom./acc.sg.neut. < **tad*).³³ The character *t* is otherwise found occasionally (for unclear reasons) in the initial position of certain verbal stems—*tbišuuant*- ‘hostile’ (= G.Av. *daibišiiant*-) < **dvišvant*- (see § 29.4.3.2.6); it is also found on very rare occasions in medial position before *k* (*aṭka*- ‘garment’ alongside *aḍka*-).

29.4.3.2.2. AVESTAN SIBILANTS

In addition to reflexes of early Iranian *s*, *š*, *z*, *ž*—including *s*, *z* < early Iranian **č*, **ǰ*—Avestan has acquired *ś* (< **čy*-)³⁴ and *ṣ* (< **rt*, see below);³⁵ in the available manuscripts the distinction that originally existed among *š*, *ś*, and *ṣ* has largely been lost, the three characters being used interchangeably. An intrusive *x* of unclear origin frequently is found in clusters of **št* and **šn* (e.g., *zixšnāgha*- ‘know’ desiderative stem < **zišnāsa*- < **jī-jnā*-).

The Iranian sequences **čv* and **ǰv* have developed into *sp* and *zb*,³⁶ respectively—cf. *aspa*- ‘horse’ < **ačva*-, *zbaia*- ‘call’ pres. stem < **ǰva*-ya-.

Voiced sibilants were often devoiced by a following nasal in Late Avestan—Gathan *uruuāzəman*- ‘joy’ = L.Av. *uruuāsman*-.

29.4.3.2.3. IRANIAN **h*

Iranian **h* is reflected by Avestan *h* in initial position and in medial position before **m*, **i*, or **u* (*haoma*- ‘ritual beverage’, *ahmī* ‘be’ 1sg.pres., *ahī* ‘be’ 2sg.pres., *vohū* ‘good’ nom./acc.neut.sg.). An *η* is introduced before medial

32. Also frequently at morpheme boundaries within a word, including regularly between the components of a compound—*druuat-biiō* ‘adherents of falsehood’ dat. pl. (= G.Av. *drəguuō-dəbiiō*).

33. “*t* ist eine sowohl tönend, als tonlos gesprochene Spirans, die dem *θ δ* nahe steht ...” (Reichelt 1978: 29). “Der Buchstabe *t* bezeichnet einen Laut, der sowohl von *t*, *d* als auch von *θ*, *δ* verschieden gewesen sein muß. Es handelt sich aller Wahrscheinlichkeit nach um einen implosiven Verschlusslaut. ... Der genaue phonetische Wert dürfte der einer stimmlosen Lenis *d* [d] ohne Plosion gewesen sein” (Hoffmann and Narten 1989: 70).

34. The palatalizing *-ii-* is normally preserved in Gathan, but tends to be lost in Late Avestan. This loss may have led to a short-lived phonemicization of *ś*, but the Late Avestan texts typically spell forms representing etymological *ś* (< **čy*-) with *š* or *ś*, indicating loss of the distinctions among these elements (Hoffmann and Narten 1989: 62–67). In some Late Avestan forms *ž* has also developed from *j* through palatalization caused by *-*y*- (*druža*- pres. stem of ‘betray’ (vs. G.Av. *adrujiiant*- ‘not deceitful’) < **družya*- < *drujya*-).

35. The development of *ṣ* < **rt* caused the compensatory lengthening of a preceding **a* (*xvāšar*- ‘drinker’ < **hvartar*-). See the discussion of accentuation below for the circumstances under which *ṣ* arose.

36. For discussion of the word for ‘tongue’ see Kuiper 1978: 12–16.

h when it is followed by **ā* (e.g., *auuaṇhō* ‘aid’ gen.sg. < **avahah*);³⁷ a similar *ṇ* developed in the sequence **hr-* > *-ṇr-*, presumably via **-ṇhr-*.³⁸ The various manifestations of **h* are in alternation with *š* < I.E. **s* after **k*, **r/ṛ*, **i/y*, **u/v* (*vīḍuṣē* ‘knowing’ dat.sg. < **viduṣ-ai* vs. nom.pl. *vīḍuāṇhō* < **vidvāh-ah*); at the boundaries of compounds *-šh-* sometimes is found instead of an expected **š*—e.g., *huš-haxā* ‘good friend’ nom.sg. rather than **hu-šaxā* < *hu-* + *haxā*—evidently resulting from an etymological spelling.

The sequences **hy* and **hv* show special phonological developments. In word-initial position they became G.Av. *ḫii*-³⁹ (L.Av. *hii-*) and *x^v-*, respectively. In medial position the treatment of these clusters differs in Gathan and Late Avestan. In Gathan the semivowels are typically treated as postconsonantal semivowels are treated elsewhere (i.e., *-hii-*, *-huu-*). In Late Avestan, in contrast, the sequences consisting of **h* + semivowel appear to have melded into a single segment (**-hy-*, **-hv-*), which, like plain **h*, became prenasalized before **ā*. The sequences spelled *-ṛjh-* and *-ṇvh-* (or *-ṇ(u)h-*) are the result.⁴⁰ Unlike Gathan, Late Avestan thus distinguishes between original **-hv-* and **-huv-*: *aojaṇhūntəm* ‘powerful’ acc.sg. (< **aujahvantam*) vs. *yāhuua* ‘in which’ fem.pl. (< rel.pron. + postpos. **yāhu ā*).

	*Iranian		Gathan	Late Avestan
* <i>hv-</i>	* <i>hvar-</i>	‘eat’	<i>x^var-</i>	<i>x^var-</i>
* <i>-hv-</i>	* <i>ahvā-</i>	‘life force’	<i>ahuuā-</i>	<i>aṇ^vhā-</i>
* <i>hy-</i>	* <i>hyāt</i>	‘be’ 3sg.opt.	<i>ḫiiāṭ</i>	<i>hiiāṭ</i> ⁴¹
* <i>-hy-</i>	* <i>vahyah</i>	‘better’ nm./ac.sg.nt.	<i>vahiiō</i>	<i>vaṇjhō</i>

Problems remain in the development of original **hu(v)-* (< I.E. **suH-*, **suu-*), which generally yields the expected *huu-* (*huuarē* ‘sun’ nom.sg., *huuaršta-* ‘good deed’) but sometimes *x^v-* (Gath. *x^vēng* ‘sun’ gen.sg. < **hu(H)an-h*), as if from **hv-*) (Beekes 1988: 15).

In word-final position **-ah* appears as *-ō* (in Gathan, also as *-ō*—see § 29.4.3.4). Word-final **-āh* takes the shape *-ā* (*mā* ‘moon’ nom.sg. < **māh*, *aməša* ‘immortals’ nom.pl. < **amrtāh*, *daēnayā* ‘religion’ gen.sg. < **dain-ayāh*). A word-final *-s* is found instead of a reflex of **-h* in sandhi with a clitic

37. The *ṇ* is frequently spread by analogy outside of its proper environment—cf. *aṇhuš* ‘being’ nom.sg. instead of **ahuš*, on the analogy of gen. *aṇhōuš*.

38. Written *-ngr-* in Gathan—cf. Gath. *angra-* ‘hostile’ vs. L.Av. *aṇra-* < **ahra-*.

39. The sequence *ḫii* < **hy* is also found in medial position in Gathan, apparently when it stood at the onset of an accented syllable—cf. *saxiiāṭ* ‘pronounce’ 3sg.opt. < **čahyāt*, *uxḍaxiiā-čā* ‘and the utterance’ gen.sg. vs. *uxḍahiiā* (Hoffmann and Narten 1989: 53–54).

40. The genitive singular ending **-ahya* typically appears as *-ahe* in Late Avestan, however.

41. On two occasions, however, *ḫiiāṭ*.

beginning with a stop—*yas-čā* relative pron. (masc.nom. sg.) + *-čā* ‘and’, *mās-ča* ‘and the moon’ nom.sg., *nəmas-tē* ‘reverence’ nom. sg. (*nəməō* < **namah-*) + *tē* ‘to you’ (< **taī*). This *-s* represents a preservation of the original In.Ir. **s*, which was blocked from the shift to Iranian **h* in obstruent clusters—see § 29.2.2 (10).

29.4.3.2.4. IRANIAN **r*/*ṛ*

Iranian syllabic **r* is represented by *ər(ə)* (e.g., *pərəsaṭ* ‘asked’ 3sg.impf. < **pṛsaṭ*); it is not clear whether this spelling is a graphic rendering of syllabic **r* or an indication that Iranian **ṛ* has developed into a “vowel-*r*” sequence.

Before the voiceless stops *p* and *k* Avestan *r* (whether of a consonantal or a syllabic origin) appears either as *-hr-* or as *-rə-*, the spelling *-hr-* presumably indicating a voiceless /*ṛ*/. In the place of an opposition **rəṭ*/*ṣ* *hrt* we find *rəṭ*/*ṣ*, indicating that **ṛ* has developed into a sibilant. The two manifestations of **r* before the voiceless stops (*rəp*, *rəṭ*, *rək* vs. *hrp*, *ṣ*, *hrk*) appear to reflect an early Iranian accentual distinction (see § 29.4.3.4).

A high vowel character (*i*, *u*) is regularly inserted before *r* in anticipation of a following high vowel or semivowel—*pouruš* ‘much’ nom.sg. < **paruṣ*, *irista-* ‘dead’ < **rista-*.

29.4.3.2.5. AVESTAN NASALS

The loss of a nasal between **ā* and a continuant, and the concomitant appearance of the nasal vowel *ã*, has been seen above, as has the shift of **a* > *ə* before a nasal (see § 29.4.3.1.3). The presence of a voiceless stop following a nasal will cause **ā* to become *ā̃* (*barānti* ‘bear’ 3pl. subjunctive, *fānkauuō* ‘peaks’ nom.pl.).

The same shift of **ā* > *ā̃* occurs before *-ŋh-*, even when the nasal is historically secondary. In historical terms, L.Av. <*-ŋh-*> represents both **-h-* (see above) and **-nh-*; Gathan too uses this spelling for both, although “there is a tendency, at any rate in some groups of manuscripts, to write *ñgh*, or *ñṇh* [i.e., <*ṇgh*>, <*ṇṇh*>—D.T.] for original *NH*, reserving *ŋh* for ancient *H*” (Morgenstierne 1973: 65). Gathan also makes a distinction between the reflexes of **-h-* and **-nh-* in that the former does not normally change a preceding **a* to *ə*, while the latter does—cf. *auuaŋhō* ‘help’ gen.sg. < **avah-ah*) vs. *māŋghāi* ‘think’ 1sg.aor. subjunctive (< **man-h-*).

It is possible that *ŋ* was phonemic, as it is found before *t* in *paŋtaŋhuua-* ‘fifth’, which results from the loss of a velar (**paŋktahva-*) (Morgenstierne 1973: 74).

Some of the manuscripts employ a character <*ñ*> to represent **n* when it is followed by *i*, *ü*, or *e*, in order to represent a palatalized realization of **n* (Hoffmann and Narten 1989: 59–62).

In L.Av. *nmāna*- ‘house’ a *d* has assimilated to an adjacent *m* to yield *nm*- (cf. G.Av. *dəmāna*-).

29.4.3.2.6. IRANIAN **v*

The Avestan reflex of the sequence **-θv*- is written as *-θβ*-, evidently indicating a shift from semivowel to obstruent. The labial in the reflex of **-dv*- may be spelled with *-β*- in Late Avestan—*aδβan*- ‘road’ (= G.Av. *aduan*-) but *vīδuuah*- ‘knowing’)—“where there is no morphological break after the voiced fricative” (Morgenstierne 1973: 61).⁴² A number of questions surround the reflex of word-initial **dv*:- for Gathan we find a distinction between **dva*- (*duuaēšah*- ‘enmity’ < **dvaišah*-) and **dvi*- (*daibišiiant*- ‘enemy’), while for Late Avestan we find a distinction between **dva*- (L.Av. *duuaža*- ‘flutter’ pres. stem) and **dv(a)i*- (*tbaēšah*- ‘enmity’, *tbišiiant*- ‘enemy’, *bitiia*- ‘second’ < **dvitiya*-).

In the L.Av. reflex of **gv* the velar is lost—*druuant*- ‘adherent of falsehood’ = G.Av. *drəguuant*-, *mourum* ‘Margiane (country name)’ acc. = O.P. *marguš*.

29.4.3.2.7. ANAPTYXIS

When followed by a high vowel or a semivowel, the consonant *r* is regularly preceded by an anticipatory vowel *i* or *u*. In addition, a vowel *a* or *ə* may be inserted before *r* when it is preceded by *s*, *z*, *f*, or *n*—*s(ə)raoša*- ‘obedience’. A vowel *ə* is also regularly inserted after *r* when it is part of a cluster *-rC*- (unless the *C* is **-y*- or **-v*-, in which case we find *-ii*- or *-uu*-) or is in word-final position⁴³—cf. *darəga*- ‘long’ < **darga*-, *vadarə* ‘weapon’ nom./acc.sg.

In addition, a *ə* is often inserted before a nasal when it is preceded by a spirant—*rafənah*- ‘support’—or, in Gathan, by *d* or *g*—*gənā* ‘woman’. Voiced obstruent clusters in Gathan also receive an inserted vowel—*aogədā* ‘said’ 3sg.

29.4.3.3. Avestan reflexes of the I.E. laryngeals

When syllabified (i.e., when found between two consonants or between a consonant and a word boundary), the I.E. “laryngeal” elements developed into the In.Ir. vowel **i* or (before **y*) **a* (see § 29.2.1 (3)). As a rule, in Iranian we find that in medial position this vowel is lost (e.g., Avestan *draonah*- ‘allotment’ vs. Sanskrit *dravinas*-). Occasionally we find in Gathan that, counter to expectations, the vowel from **H* has also been lost in the first syllable—cf. *ptā* ‘father’ nom.sg. vs. Sanskrit *pītā*.

42. The shift of Iranian **v* > *β* in these sequences seems to have blocked the shift of **-vəN* > *uN*—cf. the secondary ending 2pl.mid., G.Av. *-dum* vs. L.Av. *-δβəm* (< **-dvam*).

43. The *ə* is not inserted, however, after voiceless *f* (e.g., *vəhrka*- ‘wolf’).

Gathan meter suggests that some form of hiatus survived into early Avestan from original **H* between syllabics. For example, Kuiper (1978: 6–7) notes, “... the metre shows that the subjunctive *dāt* was still distinct from the injunctive in that the former was a disyllable [da-at], the latter monosyllabic [dāt] ...,” thus reflecting the earlier distinction between **daH-* + *-a-* + *-t* (where *-a-* was the modal marker) and **daH-* + *-t*. This suggests the existence of a hiatus in early Avestan reflecting the effects of the former laryngeal. The question of the phonological role played by the laryngeals (or by whatever developed from them) in the Avestan of Zaratustra’s time remains a matter of debate (see Kuiper 1978; Beekes 1988: 83–93).

29.4.3.4. Accentuation

Bartholomae (1886: 35–50) pointed out the link between the twin reflexes of Iranian **rt* (Avestan *š* vs. *rət*) and the location of the accent in Vedic and, it may be assumed, in early Iranian. As has been mentioned above (§ 29.4.3.2.4), the bifurcation in the development of **rt* is related to the bifurcation seen also in the sequences **rp* and **rk* (> Avestan *hrp*, *hrk* vs. *rəp*, *rək*). Some of the evidence concerning the role of the placement of the early accent in the appearance of the various reflexes of **r* + voiceless stop may be found below (Mayrhofer 1989: 12–13).⁴⁴

	<i>*Accent + rC</i>	<i>*rC + Accent</i>
<i>*_p</i>	<i>kəhrp-əm</i> ‘form’ (acc.) < <i>*kírp-am</i>	<i>*kərapəm-ča</i> ⁴⁵ ‘and the form (acc.)’ < <i>*kírp-ám-ča</i>
<i>*_t</i>	<i>aməša-</i> ‘immortal’ < <i>*a-míta-</i>	<i>aməratāt-</i> ‘immortality’ < <i>*a-mṛtātāt-</i>
<i>*_k</i>	<i>mahrka-</i> ‘destruction’ < <i>*márka-</i>	<i>marəkaē-čā</i> ‘and in death’ < <i>*markái-ča</i>

A similar indication of the role of the early accent in the phonological development of Avestan was claimed by Jackson (1891), who suggested that the two Avestan reflexes of **-auš*, the *u*-stem genitive singular ending, may be accounted for by the accent. Forms such as Avestan *vanhəuš* and *xratəuš*, with *-əuš* < **-auš*, have Indic cognates in which the accent fell upon the stem (*vásu-*, *krátu-*); in contrast, the Avestan ending *-aoš* (*mərəθiiaos*, *paraos*) typically appears when the ending of the Indic cognate was accented (*mṛtyú-*, *purú-*).

44. For a differing interpretation of the shift of **rt* > *š* see Miller (1968).

45. To be emended from *kərapəmča*.

Several features of the phonology of Gathan word-final syllables have been related by Kuryłowicz (1925) to the location of the accent. Kuryłowicz makes a convincing case for the existence of statistically significant correlations between a Vedic-like accentuational model and the G.Av. reflexes of final **-ai*, **-am/-an*, and **-ah*.

vahištē ‘best’ loc.sg. *xšaθrōi* ‘realm’ loc.sg. cf. Vedic *vásiṣṭha-*, *kṣatrā-*
 < **vāhištai* < **xšaθrái*

ahurəm ‘god’ acc.sg. *darəgēm* ‘long’ acc.sg. cf. Vedic *ásura-*, *dīrghá-*
 < **áhuram* < **dargám*

ahurō ‘god’ nom.sg. *parō* ‘excepting’ cf. Vedic *ásura-*, *parō*
 < **áhurah*⁴⁶ < **paráh*

The correlations drawn by Kuryłowicz extend beyond lexical accent to include agreement between Gathan and the grammatical placement of accent in Vedic: compare G.Av. *gərəzē* vs. *gərəzōi* ‘complain, lament’ 1sg.pres.mid. (< **grjái*) in the following passages (Kuryłowicz 1925: 5–6):

tā uxδā maniiēuš mahiiā mazdā ašāičā yūšmaibiiā gərəzē (Yasna 32, 9)
 ‘these words of my spirit, O Mazda, to you and Righteousness I
 lament (*gərəzē*).’

gərəzōi tōi ā īt auuaēnā ahurā (Yasna 46, 2)

‘I complain (*gərəzōi*) to you: look upon it, O Ahura.’

The differing reflexes of **-ai* in such passages may be interpreted by relating them to the accentuation rules governing main-clause verbs in Vedic: the verb is accented in clause-initial position (hence Iranian **grjái*) but unaccented elsewhere (hence unaccented **grjái*).

Since there is no graphic indication of the suprasegmental phonology of Avestan, it is not clear whether the various segmental manifestations of the accent which have been hypothesized are to be taken to indicate that a Vedic-like accentual system still existed in Avestan. There is, however, indirect evidence suggesting that some form of accentual shift took place when a clitic was added to a word or a derivational suffix to a stem, since the introduction of such elements occasionally leads to the shortening of an originally long vowel.

čaθβārō ‘four’ vs. *čaθβaras-ča* ‘and four’ (< **čaθvārah*)

rāsant- adj. (mng. unclear) vs. abstract *rasastāt-* (*rāsant* + noun-formant *-tāt-*)⁴⁷

46. The *ō*-reflex of final **-ah* has been generalized; the *ə*-reflex, when it occurs, largely agrees with Kuryłowicz’s analysis (1925: 12–13).

47. See also n. 39 for the role of accentuation in the appearance of medial **-xii-*.

Bibliography

- Andreas, F. C. 1904. "Die Entstehung des Awesta-Alphabets und sein ursprünglicher Lautwert." In *Verhandlungen des XIII. internationalen Orientalisten-Kongresses, Hamburg September 1902*, pp. 99–106. Leiden: Brill.
- Bartholomae, Christian. 1886. *Arische Forschungen II*. Halle: Niemeyer.
- . 1895–1901. "Awestasprache und Altpersisch." In *Grundriß der iranischen Philologie*, ed. Wilh. Geiger and Ernst Kuhn, pp. 152–248. Strassburg: Trübner.
- . 1904. *Altiranisches Wörterbuch*. Berlin: de Gruyter. Repr. 1961.
- Beekes, Robert S. P. 1988. *A Grammar of Gatha-Avestan*. Leiden: Brill.
- Brandenstein, Wilhelm, and Manfred Mayrhofer. 1964. *Handbuch des Altpersischen*. Wiesbaden: Harrassowitz.
- Collinge, N. E. 1985. *The Laws of Indo-European*. Amsterdam: Benjamins.
- Duchesne-Guillemin, Jacques. 1964. "La fixation de l'Avesta." In *Mélanges Morgenstierne 1964*: 62–66.
- Gershevitch, Ilya. 1985. "Dialect Variation in Early Persian." In *Philologia Iranica*, pp. 194–222. Wiesbaden: Reichert.
- Hertel, Johannes. 1931. "Awestisch *tb-*, *tk-*, *-tk-*." *Wiener Zeitschrift für die Kunde des Morgenlandes* 38: 91–99.
- Hoffmann, Karl. 1975. *Aufsätze zur Indoiranistik*, ed. Johanna Narten. Wiesbaden: Reichert.
- Hoffmann, Karl, and Johanna Narten. 1989. *Der sasanidische Archetypus: Untersuchungen zur Schreibung und Lautgestalt des Avestischen*. Wiesbaden: Reichert.
- Jackson, A. V. William. 1891. "The Genitive Singular of *u*-Nouns in the Avesta." *Beiträge zur Kunde der indogermanischen Sprachen* 17: 146–52.
- Kellens, Jean. 1989. "Avestique." In Schmitt 1989a: 32–55.
- . 1991. "L'avestique de 1972 à 1990." *Kratylos* 36: 1–31.
- Kent, Roland G. 1953. *Old Persian: Grammar, Texts, Lexicon* (American Oriental Series 33). New Haven: American Oriental Society.
- Kuiper, F. B. J. 1978. "On Zarathustra's Language." *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen, afd. Letterkunde* (new series) 41/4: 73–106.
- Kuryłowicz, Jerzy. 1925. *Traces de la place du ton en gathique*. Paris: Champion.
- . 1964. "L'accentuation en vieil iranien." In *Mélanges Morgenstierne 1964*: 103–7.

- Mayrhofer, Manfred. 1968. "Die Rekonstruktion des Medischen." *Anzeiger der philosophisch-historischen Klasse der Österreichischen Akademie der Wissenschaften* 1–22.
- . 1971. "Neuere Forschungen zum Altpersischen." In *Donum Indogermanicum: Festgabe für Anton Scherer*, ed. Robert Schmitt-Brandt, pp. 41–66. Heidelberg: Winter.
- . 1989. "Vorgeschichte der iranischen Sprachen: Uriranisch." In Schmitt 1989a: 5–24.
- Mélanges Morgenstierne. 1964. *Indo-Iranica: Mélanges présentés à Georg Morgenstierne à l'occasion de son soixante-dixième anniversaire*. Wiesbaden: Harrassowitz.
- Miller, D. Gary. 1968. "rt-Clusters in Avestan." *Language* 44: 274–83.
- Morgenstierne, Georg. 1973. *Irano-Dardica*. Wiesbaden: Reichert.
- Paper, Herbert H. 1956. "The Old Persian /l/ Phoneme." *Journal of the American Oriental Society* 76: 24–26.
- Payne, John R. 1992. "Iranian Languages." in *International Encyclopedia of Linguistics*, ed. William Bright, 2:228–33. New York: Oxford University Press.
- Reichelt, Hans. 1978. *Awestisches Elementarbuch*, 3rd ed. Heidelberg: Winter.
- Schmitt, Rüdiger, ed. 1989a. *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert.
- . 1989b. "Altpersisch." In Schmitt 1989a: 56–85.
- Skjærvø, Prods O. 1989. "Modern East Iranian Languages." In Schmitt 1989a: 370–83.
- Windfuhr, Gernot L. 1989. "Western Iranian Dialects." In Schmitt 1989a: 294–95.

Pahlavi Phonology

Dieter Weber

University of Göttingen

30.1. Introductory remarks

The following attempt at a phonological description of Pahlavi can only in part achieve any certainty; much must remain hypothetical and open to further discussion.¹ The aim of this article is, therefore, to present a model of the system of Middle Persian phonemes and possible allophones that fits into the historical concept of the evolution of the Persian language as a whole.

30.2. Background

30.2.1. *Language*

Pahlavi was the official language of the Sasanian Empire and, as the Empire's state religion was Zoroastrianism, of the religious and lay writings of that time. To it belong also lapidary inscriptions as well as ostraca, papyri, bullæ, and sigla. The material, therefore, is both rather voluminous and also heterogeneous.

Middle Persian as a whole continues the language of southwestern Iran where Old Persian, the official language of the Achæmenians, originated. It had also been assimilating forms and features of other dialects since Achæmenian times, but the kernel of Middle Persian is still indeed genuine Persian. Informative maps of the distribution of Middle Iranian languages are to be found in Oranskij 1960, facing p. 146, and in MacKenzie 1991: 2474.

The main features of the historical development of all MIran. languages are great changes in phonetics accompanied by simultaneous reduction or

1. AUTHOR'S NOTE: I am very grateful to Prof. Dr. D. N. MacKenzie, whose helpful instructions on many Pahlavi problems made this form of presentation possible; but of course I alone am responsible for the views presented in this article. ABBREVIATIONS: Arm., Armenian; Av., Avestan; ENP, Early New Persian; InscrMP, Inscriptional Middle Persian; JP, Jewish Persian; Khot., Khotanese (Saka); ManMP, Manichæan Middle Persian; MIran., Middle Iranian; NP, New Persian; OInd., Old Indic (Old Indo-Aryan); OIran., Old Iranian; OP, Old Persian; OstrMP, Middle Persian of the Ostraca; PapMP, Middle Persian of the Papyri etc.; Parth., Parthian; prs., present.

even loss of nominal and verbal endings. Thus there are no nominal case morphemes at all in MP except the distinction between singular and plural, where only the plural is marked (by **-ān**); grammatical relations are expressed by the use of pre- and postpositions—thus replacing the classical “cases”—or by the position of the word in the phrase. The verbal endings are reduced to one paradigm of three persons in the singular and plural for an unmarked tense (= present), while another paradigm for the same persons is in use for one marked modal form² (= subjunctive). The past is expressed by analytic forms that are built on the old past participle in conjunction with the (present) paradigm of the verb ‘to be’. It must be noted that the syntactic features used with transitive and intransitive verbs are different; for the transitive in the past the term **ERGATIVITY** is used to describe a construction in which the verb form accords with the logical object.

Phonetically, Middle Persian is characterized by retaining voiceless stops only at the beginning of a word and in voiceless surroundings (clusters); rarely, under certain and partly unknown conditions, they may occur between vowels or at the end of a word. This means in general that after vowels or in other voiced surroundings only voiced stops are possible and that, at least in part, the phonological opposition between voiced and voiceless stops is neutralized in these positions; cf. esp. Back 1981 and Pisowicz 1985. It is this phonetic development that makes it difficult in many cases to give examples for minimal pairs of stops in the above-mentioned positions.

The first phonological description that deserved this designation was given by MacKenzie 1967 and used later in MacKenzie 1971. His description is based on considerations made by Henning earlier, and he is followed now by most of the scholars of things Iranian (cf. Dresden 1970: 48ff.; Rossi 1973, 1975; Rastorgueva & Molčanova 1981; Pisowicz 1984, 1985; Sundermann 1989a, 1989b). In this paper we shall follow, on the whole, the phonological description as given by MacKenzie (with some possible modifications); for earlier descriptions of Pahlavi see Henning 1958, MacKenzie 1967, and Weber 1993: 1–3. The basis for the description will be the material discussed in § 30.2.2.

30.2.2. *Sources*

Determining the phonemes of a dead language requires more than one method and is highly dependent on the material available for the purpose. First of all there is the evidence of the (dead) language itself, (a) the **INTERNAL**

2. There are relics of another modal paradigm in some forms of the old optative.

evidence, of literary and other monuments of the language in question, its script (or scripts) with possibly different spellings; also historical aspects of that language may be of great use if the language prior to it and/or the one following it are known. Within the scope of (b) EXTERNAL evidence we have to take into account loans from that language into other (neighboring) languages and/or transcriptions of words, phrases, or whole texts in a foreign script, as well as quotations; names also play an important role in this respect.

With regard to MP we are, at first sight, in a rather good situation as there is a great number of original texts, both inscriptional and literary, as well as transcriptions of words and phrases in foreign scripts and loans in other languages. But the main crux of Pahlavi studies is the ambiguity of the script used for this language since early Sasanian times. It uses an extremely historical spelling which for a long time prevented Iranian studies from establishing the phonemic system of the language.

30.2.2.1. Internal evidence

The internal evidence must rely on the corpus of Pahlavi texts, mainly inscriptions and the bulk of Zoroastrian texts (Book Pahlavi), to a lesser extent on official and other documents (papyri, parchments, ostraca, sigla, bullæ). Since all texts are written, indiscriminately, in the same historical spelling, there are indeed very few possibilities for determining the exact pronunciation; only variations in the spelling itself may help in some cases (cf. e.g. <wpc> and <wbz> in § 30.5.3.3).

30.2.2.2. Manichæan Middle Persian

Though ManMP represents virtually the same state of Middle Persian as Pahlavi, it cannot, strictly speaking, be reckoned as internal evidence because it belongs to another religious world (Manichæism) and has another script. But it is exactly this script, derived from Semitic and adapted by the Manichæans for the special purpose of promoting their religious ideas, that gives us the clearest phonetic picture of Middle Persian. The major assistance we get from Manichæan is for the consonantism, where the voicing of intervocalic stops and affricates is clearly documented by appropriate graphic signs. With regard to the vowels there is still some uncertainty; short /a/ usually is not represented by any graphemic sign when medial but written by an aleph when initial, whereas long /ā/ is normally written by one or (less often) two alephs. With regard to other vocalic phonemes, <y>

can be used for front vowels (short or long) and <w> for back vowels (short or long); thus it is possible that <y> could stand for [i], [ī], [e], and [ē] and likewise <w> for [u], [ū], [o], and [ō]. But, fortunately, the possibilities given here are often reduced by various spellings (e.g. with or without <y> or <w>) or by historical aspects, i.e. the etymology of a word. It was Henning who first used this Manichæan writing to place Pahlavi on sounder grounds than previous scholars.³

30.2.2.3. *External evidence*

MP (Pahlavi) words may occur as loans in neighboring languages or simply be transcribed in scripts other than Pahlavi. To the last-mentioned group belong a great number of names which are therefore a very important help for establishing actual pronunciations. Here we have to distinguish between records in scripts derived from Semitic with their known insufficiencies and others with a fully developed system especially for expressing vowels. Thus, the transcriptions of a MP word, e.g. of a title or a name, in Syriac is less helpful than its rendering in Greek, Latin, Armenian, or Georgian; here, in the first place, the Greek versions of 3rd century MP inscriptions are of immense importance, e.g. Ἀδωρᾶναδ for Inscriptional MP ʾtwrʾnhyt, Γουργᾶν for InscrMP gwrkʾn, or Ὠρμισδαρταχίρ for InscrMP ʾwhrmzdʾrthštr (Back 1978: 151ff.). Armenian is replete with Iranian loanwords of Arsacid, Sasanian, and later times, so that a careful study has to be undertaken when its material is being used for MP purposes; cf. Arm. *mog* ‘magus’ ~ **mog[-mard]**; *movpet* ‘chief of the *magi*’ (Hübschmann 1897: 195) ~ **mowbed**; or *patčēn* ‘copy’ (Hübschmann 1897: 224) ~ **pa(č)čēn** (§ 30.6.2) (cf. Bolognesi 1960: 61). But in many cases one can rely also on examples from Central Asia, such as Uyghur Turkish or even Chinese; cf. the MP name *Yazdēn* transcribed by Chinese *I-yen* (Henning apud Pulleyblank 1952 :333 n.1).

30.2.2.4. *Historical aspects*

Phonemes of an intermediate stage of a language may be confirmed—or at least supported—by its historical development if earlier or later stages are known. Thus the greatest changes took place between the stage that is usually called Old Iranian (Old Persian and Avestan) and Middle Iranian: generally this is much nearer to the stage of the new Iranian languages than to the older ones, so it is justified, in the case of Middle Persian, to take into

3. Some of these, for instance Nyberg, nevertheless never changed their transcription of Pahlavi, based on the traditional spelling.

consideration the sounds of New Persian, especially its earlier “classical” form (ENP). Most of the phonological elements that are derived from Man-
MP are thus confirmed by ENP, e.g. the monophthongization of OP diph-
thongs (generally resulting in the Pahlavi phonemes /ē/ and /ō/) or the final
voicing of the **-aka*-suffix (common Pahlavi **-ag** > NP *-e*). Therefore it is jus-
tified, in the following description of the Pahlavi phonemes, to refer to
(E)NP forms very often, as is also done regularly in MacKenzie 1971. These
aspects of historical phonology may, of course, be understood as belonging
to the internal evidence, too.

30.3. Written records

30.3.1. *Script(s)*

Middle Persian of Sasanian times was written in a script that ultimately de-
rived from official Aramaic of the Achæmenian empire (see MacKenzie
1971: xi). It developed a lapidary style which was especially used in the in-
scriptions of the 3rd and later centuries but also a cursive one which was
common for everyday communication such as letters, lists, and even official
documents mainly written on leather/parchment, or ostraca, later also on
papyrus or linen (during the short occupation of Egypt in the early 7th cen-
tury). From this cursive derives the script of the Zoroastrian books; the old-
est surviving manuscripts date from the 14th century onwards. A Christian
document, the Psalter fragment (written in the 7th century and found in the
oasis of Turfan), shows some archaic and petrified forms of the script.

30.3.1.1. *Book Pahlavi*

The script of the Books is highly ambiguous. As can be seen from the list
given in MacKenzie 1971: xii, some signs denote two or even more different
sounds; thus a vertical stroke can be interpreted as *w*, *n*, *r* (or *ʾ* in ideograms),
or as a final stroke often added to the last letter of a word (especially words
ending in *-k* or *-n* and others). For *g/d/y* there is only one sign, though in later
times some diacritics similar to those of NP came into use to distinguish
them. Aleph and *h* (*x*) again have the same shape. Even worse is the situation
when combinations of the said letters occur so that double *g/d/y* could also
be read as *s*, or the combination of aleph and *g/d/y* might be interpreted as
g/d/y plus aleph etc.

It is, therefore, absolutely necessary to use a transliteration where every
letter of the Pahlavi script is rendered by a Roman character; but since the

script itself often allows various interpretations it is of no use to the reader to transcribe, for example, the vertical stroke by only one character, say by *n* or *r* or *w*. By this method the shape of a word would be totally obliterated and a resultant reading hindered. Therefore, the reader of a Pahlavi text has to decide, already when transliterating it, for one or another interpretation. Only as a second step can a transcription be made, following phonemic (perhaps partly also phonetic) rules explained in § 30.4ff.

30.3.1.2. Ideograms

Another important feature of the script is the use of Semitic (and/or quasi-Semitic) ideograms for a great number of words including nouns, pronouns, and verbs, as well as adverbs and prepositions. It is *communis opinio* (against some philologists in the 19th and at the beginning of the 20th century) that those ideograms were never pronounced in their Semitic masque but always by their MP equivalent. Some special forms of letters are mostly restricted to ideograms, e.g. <ʾYK̲> (for MP *ku*), where the <K̲> has a right-winged stroke at its foot, whereas the normal <k> has not; but, of course, there are rare instances where <K̲> is also used for [γ] in genuine Pahlavi words, e.g. **mēγ**.

For practical purposes the following conventions to distinguish (a) graphemic, (b) phonemic, and (c) phonetic rendering are observed in this article:

- (a) < > for transliteration,
- (b) bold for (phonemic) transcription, and
- (c) [] (square brackets) for phonetic spelling.

30.3.2. Specimens

Here follow specimens with (a) transliteration, (b) transcription, and (c) translation of some typical Pahlavi writings:

30.3.2.1. Early inscriptional MP

[illegible]

Inscription of *Kardīr*, *Ka'be-yi Zardošt*, ex line 2:⁴

- (a) W-KBYR 'twry ZY wlhl'n YTYBWNd W-KBYR mgwGBR'
'wlv'hmy W-ptyhwy YHWWNt W-KBYR 'twl'n W-mgwny
p'thštly HTYMWN W-'wħrmzdy W-yzd'n LB' swty YHMTWN
W-'ħlmny W-ŠDY'n LB' mħyk'ly YHWWNt
- (b) *Ud was ādur ī Wahrām nišānī ud was mowmard urwāhm ud padēx
bawēd ud was ādurān ud mowān pādixšīr āwišt ud Ōhrmazd ud
yazadān wuzurg sūd rasid ud Ahreman ud dēwān wuzurg mihgār
būd.*
- (c) And many *Wahrām*-fires (temples) were installed, and many of
the *magi* became joyful and prosperous; and for many fires
and *magi* agreements were sealed. Great profit came to
Ōhrmazd and to the gods, (but) for *Ahriman* and the *Dēws* it
became very troublesome.

30.3.2.2. *Pahlavi Psalter*

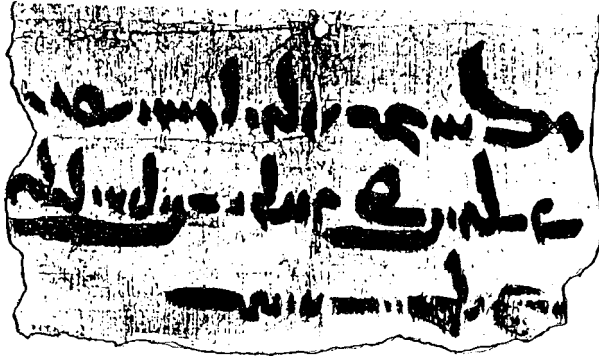
~~Handwritten text, likely bleed-through from the reverse side.~~

Fol.6, recto [Psalm 125:3] (Andreas & Barr 1933, reprinted with permission):

4. For reasons of simplification the facsimile is taken from Sprengling 1953. For reading and interpretation see now MacKenzie 1989. The text has been somewhat normalized.

- (a) ³ o MH-š L'
⁴ ws'd't špšy ZY dlwcn'n PWN
⁵ bhly ZY 'lt'dwny o WL' pl(l)'d'ndy
⁶ 'lt'y YDH PWN dlwby o
- (b) ³ o čē-š nē
⁴ *wisāyād šafš(?) ī drōzanān pad*
⁵ *bahr ī ardāyān o ud nē frārāyānd*
⁶ *ardā dast pad drō o*
- (c) ³ For not
⁴ shall rest the sceptre(?) of the wicked on
⁵ the lot of the righteous; and not shall stretch out
⁶ the righteous (their) hand(s) unto deceit.

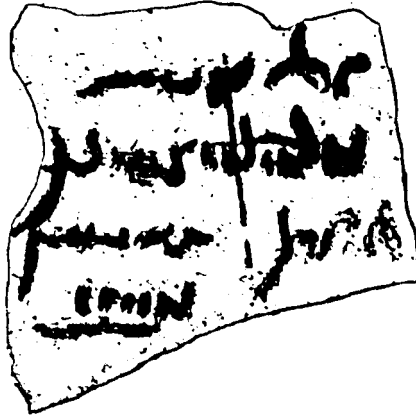
30.3.2.3. Papyrus



P.80, 7–9 (Weber 1992: 184). Egypt, ca.620–629 A.D.:

- (a) ⁷ 'L yzd'n hm'y plhw'tl kr't' pt'
⁸ ḥc'l'nwšbḥt ḥwslwybwlcynwlḥl'n
⁹ nc Y l'mšn Y pn'ḥ
- (b) ⁷ *ō yazdān hamē farroxtar kard pad*
⁸ *hazār-anōšbaxt Xusrō-Burzēn-Wahrām*
⁹ *namāz ī rāmišn ī panāh*
- (c) ⁷ To (him who) is steadily made more fortunate by the gods, (to
him who is full of)
⁸ thousandfold immortal fortune, (to) *Xusrō-Burzēn-Wahrām*
⁹ reverence, peace, protection.

30.3.2.4. Ostrakon



O.11; (Weber 1992: 16). Rayy near Teheran, presumably 7th century.

- (a) ¹ YWM gwš
² plḥw'klṭ(?) ḤS II
³ MN 'YK ḤS -/III
⁴ YḤBWNṭ'
- (b) ¹ rōz Gōš
² Farroxkard(?) may 2
³ az kū may 1/4
⁴ dād
- (c) ¹ Day Gōš (14th day).
² Farroxkard(?) wine 2 (quarts),
³ whereof wine 1/4 (quart),
⁴ given.

30.3.2.5. Sasanian Seal



(Bivar 1969, AD1, p.49, plate 3).

- (a) wḥwdynšḥpwḥly ZY 'yl'n-'nblkpty
- (b) *Vēhdēn-Šāhbūr ī Ērān-hambāragbed*
- (c) *Wēhdēn-Šāhbūr*, chief-storekeeper of Iran.

30.3.2.6. *Book Pahlavi*

from the *Ardā-Wirāz-Nāmag* (from the MS. M51b, fol.153r, lines 1–5, reprinted with permission from Hansen 1963: 52; see also *ibid.* pp. 52–53).

- (a) ¹ PWN ŠM Y yzd'n'
² 'ytnw' YMRWNd 'YK 'ywb' ḥlwb' zltwḥšt
³ ḍyn' Y MKBLWNt BYN gyḥ'n lwb'k BR' krt' 'D
⁴ bwnḍkyḥ 300 'ŠNT ḍyn' BYN 'pyckyḥ W 'NŠWT'
⁵ ḥ BYN 'pyḡwm'nyḥ YḍHWWNt ḍHWHd
- (b) ¹ *pad nām ī yazdān*
² *ēdōn gōwēnd ku: ēwbār ahlaw Zarduxšt*
³ *dēn ī padīrift andar gēhān rawāg be kard tā*
⁴ *bowandagīh <ī> 300 sāl dēn andar abēzagīh ud mardōm*
⁵ *andar abēgumānīh būd hēnd*
- (c) ¹ In the name of God.
² Thus one says: Once righteous Zarathushtra
³ made the received Religion current in the world; until
⁴ the end of 300 years the Religion was in purity and mankind
⁵ (lived) in certainty.

30.4. Phonology

30.4.1. *Phonemes*

Pahlavi includes the phonemes shown in Table 30-1.

In the main this list is in accordance with MacKenzie (1971: xiv, xv), except for /x^w/; bracketed phonemes are disputable and discussed separately.

Table 30-1. Pahlavi Phonemes

Vowels						
Short	i	(e)	a	(o)	u	
Long	ī		ā	ō	ū	
Consonants						
	p	t	χ	k		
	b	d	j	g		
	f	s	ϕ	x	h	(x _w)
		z	(ʒ)	(ɣ)		(*ɣ _w)
	m	n				
	w	r l	y			

30.4.2. Vowels

30.4.2.1. Vowel phonemes

Initially as well as medially, both short *a*, *i* and long *ā*, *ī* are possible;⁵ e.g.:

Minimal pairs of /a/ : /ā/:

I **agāh** ‘without rank’ : **āgāh** ‘aware, knowing’

M **agar** ‘if’ : **agār** ‘inactive, powerless’

Minimal pairs of /i/ : /ī/:

I ʔ⁶

M **pīh** ‘food, victuals’ : **pīh** ‘fat, tallow’

There seems to be no initial long /ū/; therefore, minimal pairs of /u/ : /ū/ are possible only medially:

M **kun** ‘do! make!’ : **kūn** ‘rump, buttocks, anus’

When final, only long vowels seem to be possible, cf.

for -**ā**: **filāsōfā** (loanword), **tā** ‘until; so that’

for -**ē**: **ayrē** ~ Av. *ayriia-* (see § 30.9.1)

for -**ī**: **sī** ‘goose’, **stī** ‘being, existence’ (Av. loanword?)

5. Examples are taken from MacKenzie 1971. I = initial, M = medial, and F = final position of the phoneme.

6. No minimal pair available; but cf. the sequence **im** ‘this’ : **u-m** ‘and by me’ vs. **ī-š** *Izāfe* (= possessive) with suffixed pronoun : **ā-š** ‘then’ with suffixed pronoun.

for -ō :	stō ‘distressed, defeated’, tō ‘thou, thee’
for -ū :	srū ‘horn, nail, prong’

There is some fluctuation with regard to monosyllables such as **nē** ‘not’, **sē** ‘three’, **ka** ‘when; if, since’, **¹bē** ‘but, out’ and others where the quantity of the vowel is difficult to determine; at least there are no minimal pairs showing that the phonematic distinction of a vowel in final position was possibly neutralized.

30.4.2.2. *The short vowels [ě] and [ō]*

From ManMP spellings it is clear that MP must have had the short vowels [e] and [o], too; MacKenzie 1967: 23f. lists some very informative examples. Also from Pahlavi spellings it is possible to think of [e] and [o] in certain circumstances:

- (a) for [e] e.g. as the result of *-y- in the following syllable in **deh** ‘country, land; village’ < *dahyu-, **keh** ‘small(er), less(er), young(er)’ < *kahyah- or **-bed** < *-pati- ‘chief’.⁷
- (b) for [o] e.g. before or after labials, cf. **borz** ‘high’, written <bwlc>; **potk** ‘sledge-hammer’, written <pwtk>; **poxtan** ‘cook, bake’, written <pwhtn>; **tom** ‘darkness’, written <t(w)m>; **-wm** (suffix forming ordinal numbers from **panjom** ‘fifth’ onwards) and the like.

Whether they had phonemic status is rather doubtful to say the least. As MacKenzie 1967: 24 rightly remarks, the “sporadic nature of their appearance in the script is only a minor argument against them.” But far more important is their later development; [e] seems to have coincided, at least in some cases, with MP /i/ > /e/ in NP, in others it reverted to /a/, and [o] obviously fell together with MP /u/ > /o/. It is not possible to ignore the examples that clearly show short [e] and [o] in Pahlavi words,⁸ but for practical purposes it is useful to accept the phoneme /a/ only and to allow three allophones, viz. [a], [e], and [o]. Unfortunately the conditions under which these allophones are to be expected are not always clear.

30.4.3. *Diphthongs*

There are no monophonemic diphthongs in Pahlavi. The sequence **-ay-** in words like **mayg**, **kayk** and others must be interpreted as being biphonemic;

7. In all these cases ManMP shows the spelling with y for the (short) vowel in the word: *dyh*, *kyh* and *-byd*.

8. Cf. the list given in MacKenzie 1967: 25.

the same holds good for *-aw-* though there is but one certain example (see § 30.4.4.6).

30.4.4. Consonantal phonemes

30.4.4.1. Stops

Stops are possible in all positions, but voiceless ones are generally restricted to initial and final positions, where they are phonemic. In intervocalic position or after (voiced) continuants such as /m/, /n/, /r/, /l/, or the like, the opposition between voiced and voiceless stops had been neutralized because Old Iranian **p*, **t*, **k* had developed into their voiced counterparts. But /p/, /t/, and /k/ were, certainly since early MP times, sometimes reinstated (1) by way of loans from other Iranian dialects or even other (non-Iranian) languages, or (2) by way of Avestan loans; and (3) /k/ (and others) was kept in positions of morphemic boundaries. These exceptions may be illustrated by the following examples:

- (1) **šakar** ‘sugar’, **yākand** ‘ruby’, **kāpūr** ‘camphor’, **nīlōpal** ‘lotus, water lily’
- (2) **Akōman** ‘Evil Thought, an arch-demon’, **ātaxš** ‘fire’, **patyārag** ‘adversary’, **rapihwin** ‘south’, **Spitāmān**; for all see § 30.9.1.
- (3) **a-kard** ‘not done, undone’, **a-pus** ‘without a son, childless’, or **pākīzag** ‘pure’ : **pāk** ‘clean, pure, holy’, **nikōh-** prs. stem ‘blame, execrate’ (preverb **ni-*).

A fourth group can be formed by some words where the retention of /k/ (or other voiceless stops or affricates) is due to the syncope of short **-a-*: **ēk** ‘one’ < **aiw(a)ka-*, **nēk** ‘good, beautiful’ < **naiw(a)ka-*, **pāk** ‘clean, pure, holy’ < **pāw(a)ka-* or **ēč** with a negative ‘not any’ < **aiw(a)-čit*. Voiceless stops or affricates may also be retained when they are in secondary contact with other voiceless sounds so that the result is gemination, cf. **pattūg** (§ 30.6.2).

The reason for retained /-k-/ in the words **čakōk** ‘lark’ and **čakuč** ‘hammer’ is not clear; it may be that semantic reasons played a certain role here, as is also to be seen in the diminutive suffix /-ak/.

30.4.4.1.1. LABIAL STOPS

Minimal pairs of /p/ : /b/:

- | | |
|---|--|
| I | pīm ‘pain, pang, sorrow’ : bīm ‘fear, terror’ |
| | pand ‘path; counsel, advice’ : band ‘bond, link’ |
| F | not possible |

30.4.4.1.2. DENTAL STOPS

Minimal pairs of /t/ : /d/:

- I **tās** ‘die (for gaming)’ : **dās** ‘sickle’
tazīdan ‘run, flow’ : **dazīdan** ‘burn, scorch’
- F **ast** ‘is’ : **azd** ‘known’

30.4.4.1.3. VELAR STOPS

Minimal pairs of /k/ : /g/:

- I **kardan** ‘do, make, act, perform’ : **gardan** ‘neck’
kām ‘will, desire, purpose’ : **gām** ‘pace, step, stride; a yard’
- F **sāk** ‘tribute’ : **sāg** ‘number’
-ak (diminutive suffix) : **-ag** (suffix forming agent nouns from present stems)
zardag ‘yolk’ : **zardak** ‘safflower’

30.4.4.2. Continuants

All continuants (sibilants, nasals, fricatives as well as /r/ and /l/) are possible in all positions; for clusters formed by them see § 30.5.

30.4.4.2.1. SIBILANTS

Minimal pairs of /s/ : /š/:

- I **sēb** ‘apple’ : **šēb** ‘declivity’
sāmān ‘limit, boundary’ : **šāmān** pl. of **šām** ‘evening meal, supper’
- M **ōsān** ‘throw down!’ : **ōšān** pl. of **ōš**
- F **dōs** ‘gum; plaster’ : ¹**dōš** ‘shoulder’, ²**dōš** ‘last night, previous evening’
kas ‘person, somebody’ : **kaš** ‘armpit’

Minimal pairs of /s/ : /z/:

- I **sang** ‘stone, rock’ : **zang** ‘ankle, shank’
sūr ‘strong; meal, feast, banquet’ : **zūr** ‘false, deceitful’
- M **šnās-** prs. stem ‘know, recognize’ : **šnāz-** prs. stem ‘swim’
- F **wars** ‘hair’ : **warz** 1. ‘work, agriculture’, 2. ‘miracle’, 3. ‘club, mace’

30.4.4.2.2. NASALS

Minimal pairs of /m/ : /n/:

- I **mang** ‘henbane’ : **nang** ‘shame, modesty’
mān ‘house, dwelling’ : **nān** ‘bread, food’

- M **zamān** ‘time, hour’ : **zanān**, pl. of **zan** ‘woman, wife’
 F **rām** ‘peace’ : **rān** ‘thigh’

30.4.4.3. *Affricates*

Minimal pairs of /č/ : /j/:

- I **čār** ‘means, remedy’ : **jār** ‘time, occasion’
 M not known
 F not known⁹

Final /č/ is rare and restricted (a) to some monosyllabic words of pronominal derivation:

- ēč** (with negative) ‘not any’
kāč ka ‘could that, if only’

and (b) to some words of unknown origin or dialectal variation:

- baškuč** ‘gryphon’
gač ‘plaster, gypsum’
kafč ‘ladle, skimmer’ (see above)
čakuč ‘hammer’ (: NP *čakuš*)

30.4.4.4. *Absence of ž*

Though there are the two correlative phonemes /s/ and /š/ there are no such on the voiced level; at any rate it seems useful to give some occurrences of /z/ : /j/ (see above):

- I **zādag** ‘born; child’ : **jādag** 1. ‘omen’ 2. ‘form, property’
 zang ‘ankle, shank’ : **jang** ‘struggle, battle, fight’
 M not known
 F **bāz** ‘hawk, falcon’ : **bāj** ‘tribute, tax’

30.4.4.5. *The phonemes /r/ and /l/*

In OP /l/ is found only in three foreign names; in MP, however, an /l/ has developed from the old cluster **rd*. There are of course also some loans with foreign *l*. Furthermore, it may be possible that some words have an inherited **l* although this cannot be proved with certainty because of the scarcity of the OP material. There are only a few words in Pahlavi with initial /l-/; thus it is impossible to give minimal pairs for /r-/ : /l-/; but cf. e.g. **lab** ‘lip’ : **nab** ‘grandson’ : **šab** ‘night’.

9. But cf. **kafč** ‘ladle, skimmer’ : **kafš** ‘shoe’.

- M **wāridan** ‘rain’ : **wālidan** ‘grow, increase, prosper’
bārag ‘mount, horse’ : **bālag** ‘wing’
 F **pīr** ‘old, aged, ancient’ : **pīl** ‘elephant’¹⁰

It must be noted that in ManMP we often find, under certain circumstances, *r* instead of *l*, e.g. *sʾrʾr* for **sālār** or *kmʾr* for **kamāl** ‘head (*daēvic*)’.

30.4.4.6. *The phonemes /w/ and /y/*

Both phonemes can be treated like other continuants, e.g. /m/, /n/, /r/, /l/, /s/, or /ʃ/. They are met with in initial, medial, and final position but have a restricted occurrence in clusters (see § 30.5).

Minimal pairs for /w/:

- I **warr** ‘wool; garment’ : **parr** ‘feather, wing’
wēz- prs. stem 1. ‘sift, select, choose etc.’, 2. ‘swing, brandish’ :
mēz- prs. stem 1. ‘suck’, 2. **mēz(išn)** ‘blinking, twinkling’
 M **arwand** ‘swift, valiant’ : **aryand** ‘horrible’
zarwān ‘old age’ : **zarmān** ‘old man; old age, decrepitude’
parwāz ‘flight’ : **pardāz-** prs. stem ‘accomplish’
 F **daw(iđan)** prs. stem 1. ‘run’, 2. ‘speak’ (*daēvic*, evil) : Day ‘Cre-
 ator; 10th month, 8th, 15th, 23rd days’
nēw ‘good, brave’ : **nēm** ‘half’
marw ‘herb’ : **mard** ‘man’

Minimal pairs for /y/:¹¹

- I **yask** ‘illness, sickness’ : **nask** ‘a division of the Avesta’ (see
 § 30.9.1)
yōz ‘cheetah, hunting leopard’ : **gōz** ‘walnut’ (both with non-
 Persian -z)
yašt(an) past stem ‘worship, celebrate, recite’ : **dašt** ‘plain, open
 ground’ : **hašt** ‘eight’
 M **māyišn** ‘copulation’ : **mārišn** ‘perception’
mayg ‘locust’ : **marg** ‘death’
kayk ‘flea’ : **kark** ‘chicken, hen’

10. Cf. also **dil** ‘heart’ : **diž** ‘fortress’.

11. Here must be noted that initial /y-/ is of non-Persian origin, as OIran. *y- regularly changed to Persian *j-*; in these cases initial /y-/ must be due to Avestan (cf. especially **yasn** : **jašn**) or some other dialectal influence (cf. Parth. **yad**).

- F **ōy** (dem. pron.) : **ōh** ‘so, thus’
 nāy ‘tube, flute, clarion’ : **nān** ‘bread’ : **nāf** ‘family’

Neither phoneme forms diphthongs with a preceding vowel, especially /a/. Thus, as stated above, words like **mayg** ‘locust’ (written <mdk>), **payg** ‘foot-soldier, courier’ (written <pdk>)¹² or **kayk** ‘flea’ (written <kdyk>) are phonemically equivalent to **mazg** ‘brain, marrow’, **marg** ‘death’ and others. This may also be shown by the sequence **bay** ‘god, lord’ : **bar** ‘fruit, produce, profit’ : **pay** ‘foot’ : **pas** ‘then, afterwards’; **bay** ‘god, lord’ : **bōy** ‘smell, scent’ : **bōr** ‘reddish-brown’ : **gōy** ‘ball’ : **gōr** ‘wild ass’. From the behavior of /y/ in these instances we may infer that /w/ can also be treated in the same way, though there are only a few examples illustrating it, e.g. **lawz(ēnag)** ‘almond sweetmeat, marzipan’ : **larz(īdan)** prs. ‘shiver, tremble’.

30.4.4.7. *The phonemes /h/ and /x/*

Both phonemes are to be found in initial, medial, and final positions but their phonemic character can clearly be shown only initially:

- hast** (= **ast**) ‘is’ : **xast(an)** ‘wound, injure’
har (= **harw**) ‘all, each, every’ : **xar** ‘donkey, ass’¹³

In intervocalic position /h/ is common (cf. **dahān** ‘mouth’, **mēhan** ‘home’, **wahār** ‘spring’), whereas /x/ seems to be possible only at morpheme boundaries (**mēx(ān)** pl. ‘peg, nail’). In medial position /x/ is regularly met with in clusters where the second element is a voiceless stop (**baxtan** ‘apportion, distribute’, **juxt** ‘pair, couple’) or a voiceless continuant (**baxšišn** ‘bestowal, distribution’); here /h/ is impossible. In combination with *r* or *l* both phonemes are possible (cf. **suxr** ‘red’, NP **surx**, **taxl** ‘bitter’, NP **talx** but ManMP *thr*) as well as before nasals (**dahm** ‘virtuous, pious’ : **daxmag** ‘tomb’, but cf. also **tōhmag** ‘seed; stock, family’, but NP *tuxma*). These examples indicate that in medial position the phonemic distinction between /h/ and /x/ was perhaps irrelevant; at any rate the fluctuation between *h* and *x* in Pahlavi, ManMP, and NP seems to support this interpretation.

In final position both /h/ and /x/ are possible, cf. **rōbāh** ‘fox’, **sūrāx** ‘hole’ (NP idem). Whether here, too, the phonemic distinction between /h/ and /x/ is abandoned seems doubtful since there are pairs like **tah** ‘bottom’ : **tag** ‘running, attack’ or **mēx** ‘peg, nail’ : **mēy** ‘cloud, mist’ (see § 30.4.4.8). For the distinction between /x/ and /x^w/ (cf. e.g. **frāx** vs. **frāx^w**) see § 30.4.4.9.

12. Possibly also **gayg** (MacKenzie 1971 **gēg**) ‘thief’ (written <gdk>), ManMP *gyyg*?

13. Cf. also **xarān** ‘donkeys, asses’ : **garān** ‘heavy, serious, difficult’ : **karān** ‘side, edge, end’.

30.4.4.8. *The phoneme /ɣ/*

The phonemic character of γ in Pahlavi is disputable. There is only a small number of words containing γ in Pahlavi, and the following list repeats those given in MacKenzie 1971. The general view, however, still holds that this phoneme, if it is one, occurs “only in non-Persian, i.e. ‘learned’ or loaned, words” (MacKenzie 1971: xv).

ayrē <’k-, ’g-> (Av. *agriia-*) ‘highest, greatest’

ahlomōγ <’hlwmwk’> ‘heretic’

arγand <’lgnd> ‘horrible’, cf. NP ~

bāγ <b’g> ‘garden, orchard’

bayōbaxt (Av. *bayō.baxta-*)¹⁴ ‘fortune dispensed by the gods, providence’

čirāγ <cl’γ> ‘lamp’, NP ~

Parth. **darāγ** <dl’g> ‘bell’, cf. JP ~, but NP *darā*.

γarrānāg <gl’n’k’> ‘thunder’, **γarrānīdan**, **γarrān-** <gl’n’yt’n’> ‘roar, thunder’, **-išn** ‘roaring’, cf. NP *γurrīdan*

***jaγ** <y’h> ‘sissoo tree’, cf. NP ~, OP *yaka-*

juγ <ywg> ‘yoke’

maγ <mγ> ‘hole, pit’, cf. Av. *mayā-*

muγ [written ideographically!] ‘date-palm’, cf. NP *mux*

raγ <lgγ> ‘quick, swift’

rāγ <l’g> ‘meadow, plain’

rōγn <lwkn’> ‘butter’, cf. Parth. *rwgn*, NP *rauγan* (and ManMP *rwyyn*).

stēγ <stytk’> ‘erect’, cf. NP *sitēγ*¹⁵

waydān <wyt’n’> = Av. *vaγdana-* ‘(ahuric, evil) head’

warāγ <wl’g> ‘crow’, cf. JP *br’γ*

As can be seen there is practically only one word with initial γ -, viz. *γarrānāg* <gl’n’k’> ‘thunder’, *γarrānīdan*, *γarrān-* <gl’n’yt’n’> ‘roar, thunder’, *-išn* ‘roaring’, which may be a dialectal word. So we are left with the other instances where the [ɣ] is always in postvocalic position. But here it must be observed that a number of the examples are mere transcriptions—more or less correct—of Avestan words, so that they should not be taken into consideration with regard to the question of the phonemic status of [ɣ]. The other examples show one common feature, viz. the phoneme in question being written with <g>, <k>, or <h> and in rare cases also <γ>, which must be

14. A dot in an Avestan word, as found in the original, generally indicates a compound.

15. For phonetic reasons cf. also **tēx** <tyh> ‘sharp edge, ridge; ray’: NP *tēγ*.

interpreted as different ways of expressing the sound [ɣ]. There are some minimal pairs that show, with certainty, the phonemic status of /ɣ/:

mēy <myɣ> ‘cloud, mist’ vs. **mēx** <myh> ‘peg, nail’, cf. Parth. *myx*, NP ~ **wazay** <wzɣ> ‘frog’, cf. NP *wazay*, *bazay* vs. **wazag** agent noun of **wazidan** ‘move, blow (of wind)’.
ray <lgɣ> ‘quick, swift’ vs. **rag** <lk> ‘vein’, NP ~

Furthermore, in most cases MP *g* < **k* is kept distinct from old **g* through NP; this, in fact, is an additional argument in favor of the phoneme /ɣ/. Therefore it does not seem wise to follow Rossi, who denies /ɣ/ must not as a phoneme for “classical” Pahlavi.¹⁶ However, dialectal words may have penetrated rather early, so that a phoneme /ɣ/ could have been established; the example of Parth. *darāy* ‘bell’ (from the *Draxt-ī Asūrīg*) may be taken as a hint for the door by which dialectal words could have penetrated. This would be in conformity with the generally accepted view that northwestern elements had been entering “classical” Pahlavi since Arsacid times (Henning 1958: 92ff.). This fact and the intrusion of Arabic words with /ɣ/ from the 7th and 9th centuries onwards caused the existence of the phoneme /ɣ/ in New Persian.

30.4.4.9. The phoneme /xʷ/

This phoneme is rare and not generally accepted. It has been discussed neither by MacKenzie 1971 nor by Rossi. The latest discussion of Middle Persian phonetics (Sundermann 1989b) only mentions the fact that some authors prefer both /x/ and /xʷ/. This view was promoted by Rastorgueva and Molčanova 1981, and hesitatingly by Lazard 1963, but is strongly supported by this author. The reasons for assuming a phoneme /xʷ/ are given in Weber 1994, but some may be repeated here to give a rough idea of the problems involved.

Unfortunately the examples for a phoneme /xʷ/ are few; for medial position as well as final position it seems to be impossible to give minimal pairs, so that only in initial position can this phoneme be established with some certainty; cf. the following examples:

xʷad, also **xrad**, **xad**, and **wad**, not **swad*, **srad*,¹⁷ **dwad* and others
xʷān, also **xān**, **wān**, not **swān*, **srān*, **dwān* and others.
xʷār, also **xār** and **wār**, not **swār*, **srār*¹⁸

16. See also Pisowicz 1985: 153, who does not list /ɣ/ as a phoneme.

17. According to Tafazzoli 1970: 87–89, **srat** ‘street’ is to be read **sard** ‘échelle’.

18. But **dwār**- is possible, see § 30.5.1.1.

Obviously the situation is to be compared to that of /tʃ/ (= /č/), see **čarm** : **šarm** : **tarm(ānih)**, where **twarm*, **tmarm*, etc. and **kšarm*, **pšarm*, etc. are impossible. The phoneme /č/, however, gets some support through the existence of its voiced counterpart /dʒ/ (= /ǰ/); but both are primarily to be found in initial position.¹⁹

It must be stressed that reasons of historical phonology also play an important role insofar as a general tendency can be observed with regard to consonantal clusters of the type Old Iranian **Cw* which, within Persian, are monophonemicized: OIran. **dw-* becomes *d-* as in **dwar(a)-* ‘door’ > **dar**, OIran. **θw* becomes *h* as in **piθβ-* (some oblique case of **pitu-* ‘meat’) > **pih** (written <pihw’>) ‘food, victuals’, OIran. **św* becomes *s* as in **aśwa-* ‘horse’ > OP *asa-*, etc. From this tendency it may be inferred that OIran. **hw* became *x^w* and simply *x* later on in NP. In the main this is the view promoted by Lazard and Pisowicz. It may be added that if there was a phoneme /x^w-, theoretically the voiced counterpart /ɣ^w-/ could have arisen, which in the course of time gave /g(u)-/ or /b-/ according to the following vowel; cf. Salemann 1895: 269 and Back 1978: 131f. Thus /x^w/, and theoretically also /ɣ^w/, are necessary when regarding the genesis of Middle Persian in a structuralist way.

30.4.5. Allophonic rules

30.4.5.1. Vowels

One main feature of MP vocalism is the possible fluctuation in pronunciation of the phoneme (short) /a/. As was discussed earlier (see § 30.4.2.2), in labial surroundings /a/ could have a labialized value, viz. [o], in palatal or guttural ones a palatal value like [e]; the fact that these pronunciations are sometimes expressed in the script and sometimes not may point to the allophonic status of both for the phoneme /a/. But here in certain cases the ManMP spelling is more consistent, and it is possible that in the course of time the allophones may have been phonemicized. For all other vocalic phonemes there are no hints of the existence of allophones.

30.4.5.2. Consonants

The voiced stops will have had spirant allophones in voiced surroundings except after nasals; cf. **widaštan** ~ NP *guḏaštan/guḏaštan*. With regard to the rules of medial and final clusters as established in § 30.5.2 and § 30.5.3, the

19. Both phonemes are sometimes reinstated secondarily in medial or final positions, cf. **gač** ‘plaster, gypsum’ or **bāj** ‘tribute, tax’ (but dialectally also **bāz**); see § 30.4.4.3.

(voiced) stops in the sequences /-bR(-)/ (cf. **gabr**), /gR(-)/ (cf. **dagr**), or */-dR(-)/²⁰ and the first one in /-bd(-)/ (cf. **abd(ih)**) possibly had a spirant pronunciation [-βR(-)], [-ɣR(-)], or [-ðR(-)] (as in [aβd(ih)]), though NP still shows [-br], etc.; see also § 30.5.3.3. The same holds for words like **abzōn** ‘increase, growth’, **abzūdan** ‘increase, grow’, or **abrōxtan** ‘light, kindle, illuminate, polish’, which could have been pronounced [aβzōn], [aβzūdan], and [aβrōxtan].

Nasals are always voiced. There exists an old distribution of /m/ and /n/ with regard to the following consonant: /m/ is only accepted before labials, /n/ only before dentals, gutturals, sibilants, and affricates. Exceptions are possible under certain circumstances, e.g. morpheme boundaries; cf. **pam-bag** ‘cotton’, but **pand** ‘path; counsel, advice’, **panj** ‘five’, **pānzdah** ‘fifteen’; before gutturals we must accept only /n/ though here we have to accept the allophone [ŋ], cf. **mang** ‘henbane’, **nang** ‘shame, modesty’, and before /č/ and /j/ an allophone [ɲ] was likely.

30.5. Clusters

Consonant clusters are met with in Pahlavi quite often, but the number of possible combinations is restricted, as is the position in the word (initial, medial, or final) in which they can appear; at morpheme boundaries special rules apply.

30.5.1. Initial clusters

Initial clusters are restricted to the combination of a stop (C) or a continuant (R) + *r*, *y*, or *w* (§ 30.5.1.1) and to some combinations with initial /s-/ (§ 30.5.1.2) or /š-/ (§ 30.5.1.3).

30.5.1.1. Initial clusters of the types Cr- and Rr-

- br-** **brād** ‘brother’, **brāh** ‘brilliance, splendor’, **brāzīdan** ‘shine, gleam’, **brīštan** ‘roast’, **brūg** ‘(eye)brow’
- dr-** **drafš** ‘banner’, **drāyīdan** ‘speak (*daēvic*), chatter’, **drōd** ‘health, well-being’, **društ** ‘harsh, rough, severe’

20. There is no actual occurrence of this cluster, which changed into /-hr(-)/ as can be seen by *muhr* ‘seal’ < **mudra*-; this word at the same time corroborates the above interpretation because it must have had a pronunciation [muðra-] since it reveals, in principle, the same northwestern development as that of Iranian *-θr- > /-hr-/ (cf. **puθra*- ‘son’ > Parth. *puhr*).

- fr-** **fradāg** ‘tomorrow’, **frasang** ‘parasang’, **frāy** ‘more, much’, **frēb** ‘deceit, deception’, **frōg** ‘brilliance, brightness’, **frušag** ‘bee-stings’
- gr-** **grāmag** ‘possessions, wealth’, **griftār** ‘held, taken; captive’, **grōh** ‘group, crowd’
- sr-** **sraw** ‘word, spell’, **srištan** ‘mix, knead’, **srōd** ‘song’, **srū** ‘horn, nail, pong’
- tr-** **tru(f)š** ‘sour’, **truftag** ‘stolen; intercalary (days)’
- xr-** **xrad** ‘wisdom, reason’, **xrīdan** ‘buy’, **xrōs** ‘cock’
- zr-** **zraw-** prs. (= **azruftan**, **azraw-**) ‘go out, be extinguished’, **zrēh** 1. ‘sea, lake’, 2. ‘armor’

Examples for the clusters Cy- and Cw- (or Ry- and Rw-) are very few; for the former we may quote **gyāg** ‘place’, **gyān** ‘soul, ghost’, **syā** ‘black’, and **zyān** ‘loss, harm, damage’. But whether a monosyllabic or bisyllabic reading, viz. **siyā** and **ziyān**, should be preferred is a matter of etymology.²¹ At any rate a bisyllabic reading would be parallel to the spelling of the words **ruwān**, viz. <lwbn’>, ‘soul’, or **zuwān**, viz. <zwb’n’>, ‘tongue, language’ (= **uzwān**).

For the possible cluster Cw- (Rw-) only three examples can be adduced, viz. **dwāridan** ‘run, move’ (*daēvic*), **dwāzdah** ‘twelve’, and **swah-** prs. ‘sigh’. As it was already underlined that MP originally had no cluster /dw-/ because OIran. **dw-* became *d-* as in **dar** ‘door’ < **dwar(a)-*, words with the cluster /dw-/ must be secondary; as a matter of fact, **dwār-** seems to be a loan from Avestan, and for **dwāzdah** an initial **duwā-* must be postulated. With regard to the prs. stem **swah-**, it could have originated in a dialect form.²²

30.5.1.2. Clusters with initial /s-/

Clusters with initial /s-/ are restricted to combinations with /-n-/ , /-p-/ , and /-t-/; those with initial /sr-/ , /sw-/ , and /sy-/ are dealt with in § 30.5.1. Since OIran. **sk-* became MP **šk-**, there is no indigenous word with initial **sk-** to be found in Pahlavi; ***Skandar** as a popular form of *Alexander* (cf. NP *Iskandar*, *Sikandar*) is a loan from the Greek (see § 30.9.2). Of special interest is the cluster **sn-**, as generally IE **s* had become *h* in Iranian but obviously was kept in some special combinations; *s* seems to have kept its voiceless charac-

21. By the way, the word for ‘young woman’, written <zyd’nk’>, is given in MacKenzie 1971 as **ziyānag**.

22. The connection with Old Indic *śvas-*, as given by Henning in Boyce 1954 s.v., is not certain, to say the least, as one would expect **spah-* for Parthian and **sah-* for Persian.

ter, though other Iranian languages sometimes show a voicing of *s*- because of the voiced character of the following *n*, e.g. Khotanese *ysānāh*- ‘swim’ < **snāfyā*-; see also **šnāz**- under § 30.5.3). Unfortunately, no examples can be presented for initial /sm-/ , though **s* ought to be kept, as in the cluster /sn-/ , but cf. the name *Samarkand* ~ Μαρκανδα (which, of course, is no evidence for Persian in its earlier stages).

snāh ‘blow, strike, violence’, **snēzag** ‘snow’
spar ‘shield’, **spēd** ‘white’, **spōz** ‘disobedience; delay’, **spurz** ‘spleen’²³
stabr ‘big, coarse, strong’, **stēzag** ‘quarrel, strife’, **stūdan** ‘praise’

30.5.1.3. Clusters with initial /š/

Clusters with initial /š-/ are known from the combinations /šk-/ and /šn-/ , but it must be noted that ManMP and Parth. show a short prothetic vowel before these clusters, generally expressed by ‘Ayin; cf.

škarag ‘(bird) of prey’, cf. Parth. prs. stem ‘škr-
škēbāg ‘patient’, cf. Parth. prs. stem ‘škyb-
¹**škōh** ‘poor’, ManMP ‘škw
šnāxtan ‘know, recognize’, ManMP ‘šn’xtn
šnāz- prs. ‘swim’,²⁴ ManMP prs. ‘šn’z-
šnōšag ‘sneeze’
šnūg ‘knee’, ManMP ‘šnwg

30.5.2. Internal clusters

Generally, internal clusters are formed by a continuant followed by a stop:

-mC- **čambar** ‘circle, hoop’, **pambag** ‘cotton’, **stambag** ‘oppressive; obstinate’; **šamšēr** ‘sword’, **šimšār** ‘box tree’
-nC- **bandag** ‘servant’, **zangīg** ‘negro’
-rC- **barbut** ‘lyre’, **čarbišn** ‘fat, oil’, **murdār** ‘carrion’, **pardag** ‘veil, curtain’, **zargōn** ‘golden, green, fresh’, **Gurgān** ‘(the province) Gurgān’;²⁵ **darmān** ‘remedy, medicine’, **murwāg** ‘omen’
-sC- **bōyestān** ‘(flower) garden’, **dōstīh** ‘friendship, love’, **mastarg** ‘skull’, **pistag** ‘pistachio nut’, **rāspīg** ‘second officiating priest’
-zC- **mazgīt** ‘mosque’ (loanword), **nōzdah** ‘nineteen’, **pazdagīh** ‘expulsion’, **sēzdah** ‘thirteen’

23. Also possible /spr-/ in **sprahm(ag)** ‘flower, fragrant herb’.

24. Cf. Khot. *ysānāj*- prs. ‘bathe’ Emmerick 1968: 113, < **snāčaya*-.

25. Not in Henning 1971, only adj. **gurgānīg**.

- šC- **laškar** ‘army’, **rēškēn** ‘hurtful, harmful’, **taškanag** ‘under-skirt’,
Tištār ‘(the star) Sirius’, **waštīh** ‘sickness, illness’
- xC- **poxtan** ‘cook, bake’, **šnāxtan** ‘know, recognize’
- fC- **raftan** ‘go, move, proceed’, **škuftan** ‘flower, blossom’; **šafšēr** =
šamšēr

Words like **yōjdahr** ‘holy, pure’ can be treated here, too, if the affricate phoneme had a variant [ž] before (voiced) stops.

There are a number of internal clusters formed of two continuants (cf. **ēwarzīg** ‘moving’); they behave like those in final position (see § 30.5.3.2). Disputable cases like **abdīh** ‘wonder’, **dagrand** ‘long’ (> *dērand*), **gubrāg** ‘awake, alert, vigilant’, or **stabrag** ‘shot silk’ are treated under the corresponding final clusters (see § 30.5.3.3).

30.5.3. Final clusters

Here clusters consisting of more than two elements are not allowed. One or both of these elements must be a continuant, but cases where the first element is a stop are rare and are difficult to explain phonologically.

30.5.3.1. Examples with -RC

- mC **sumb** ‘hoof’, **xumb** ‘jug, pitcher’
- nC **band** ‘bond, link’, **druwand** ‘evil, sinful, unrighteous’, **tund**
‘sharp, violent’; **bang** ‘henbane’, **pīring** ‘monochrome damask’
- rC **čarb** ‘mild, gentle; fatty, oily’, **mard** ‘man’, **warg** ‘leaf’
- sC **dast** ‘hand’, **frasp** ‘roof-tree, beam, mast’, **nask** ‘a division of the
Avesta’²⁶
- zC **mizd** ‘reward, hire, pay’, **nazd** ‘near’, **pazd** ‘oppression, persecu-
tion’, **ruzd** ‘greedy’; **spazg** ‘slandorous’
- šC **dašt** ‘plain, open ground’, **drubušt** ‘fortress’; **kašk** ‘dried butter-
milk’, **srešk** ‘drop, tear’, **hušk** ‘dry’
- xC **baxt** ‘fortune, fate’, **passāxt** ‘test’
- fC **naft** ‘moist, damp; naphtha’, **staft** ‘hard, firm; harsh, tyrannous;
very’

For examples of the type -wC and -yC see § 30.4.4.6.

26. Though this is a *mot savant* because of the retained -sk (see § 30.5.1.2), it is Pahlavi nevertheless.

30.5.3.2. *Examples with -RR*

- mR **x^wamn** ‘sleep, dream’
- nR **aspinj** ‘hospitality; inn’, **panj** ‘five’
- rR **arz** ‘worth, value’, **ars** ‘tear(drop)’, **sarw** ‘cypress’
- sR **yasn** ‘adoration, worship’, **ēsm** = **ēzm**
- zR **bazm** ‘meal, feast’, **ēzm** ‘firewood, fuel’
- šR **arešn** ‘cubit’, **āyišm** ‘moon’, **pašm** ‘wool’
- xR **maxš** ‘fly’, **taxl** ‘bitter’, **zaxm** ‘wound’
- hR **sprahm** ‘flower, fragrant herb’, **gahl** ‘sorghum, giant millet’,
zōhr ‘libation, offering’
- fR **karafs** ‘celery’, **wafr** ‘snow’, **zofr** ‘deep’

To this group belongs also **rōγn** ‘oil, butter’ (with the questionable phoneme /γ/) unless it be regarded phonemically as */rōgn/ and treated like other words discussed in the following paragraph under (b).

30.5.3.3. *Clusters with initial stop*

From the examples given above it seems to be a general rule that in clusters only the second element can be a stop, be it voiced or voiceless. But there are some rare and disputable, but nevertheless striking, cases of the type -CC- (a) and -CR- (b).

(a) Here can be cited only

- abd** ‘wonderful’, written <’pd>
- kabk** ‘partridge’, written <kpk’>
- potk** ‘sledge-hammer’, written <pwtk’>

It must be remarked that **abd** is written in ManMP *’βd-* with a special graph for denoting the spirant -β- which is met with rarely in ManMP texts but regularly in Manichæan Sogdian texts. Both **kabk** and **potk** live on in NP.

(b) -RR-

- babr** ‘tiger’, written <bpl>, NP ~
- dabr** 1. ‘sinner, deceiver’, 2. ‘*dark-colored’, both written <dpl>
- gabr** ‘hollow, cavity, womb’, written <gpl> and <gwbl>, which rather indicates a reading [gawr], NP *gōr*
- stabr** ‘big, coarse, strong’, written <stpl>, NP *sitabr*
- dagr** ‘long’, written <dgl>, NP *dēr*
- sagr** ‘sated, satiate’, written <sgl>, NP *sēr*
- šagr** ‘lion’, written <šgl>,²⁷ NP *šēr*

Here the words with written <-g-> before the final continuant /-r/ develop long /ē/ out of /a/ plus /g/; those with written <-p-> before the final /-r/, however, retain the stop in NP. The development **-agr* > *-ēr* is supposed to have gone via an intermediate stage of *-aγr* like that of **tigr** ‘arrow’ (written <tgl>) > **tiγr* > NP *tīr*. But since there is no indication of a phonemic status of [γ] in the words mentioned, it seems adequate to assume the phoneme /g/ in these clusters; on the other hand, there is some parallelism between the cases where *-agr* developed into *-ēr* and *gabr* because of the fact that *-g-* and *-b-* have their position by metathesis from older **darga-* ‘long’, **šarga-* ‘lion’²⁸ or **garba-* ‘womb’. Thus it may be asked whether the metathesis produced a different phonological pattern than in the case of e.g. **stabr** < **stab-ra-* (root **stamb-*).

Here can be added the words **wabz** ‘wasp’ (written <wpc>²⁹ and <wbz>) and **sabz** ‘green, fresh’ (written <spz> and <sbz>); the question whether the actual pronunciation was [waβz] or [saβz] or simply [wabz] or [sabz] must remain open to speculation.

30.5.3.4. Three-consonant clusters

Clusters with more than two consonantal elements are possible but obviously only at morpheme boundaries, cf. **Ardwahišt**, **mihrbān** ‘friendly, kind’, **murwniš** ‘augur, soothsayer’, **waxšwar** ‘prophet’, or **yōzdahrgar** ‘purifier’, where in most cases the nominal compounding is obvious. Therefore, a word like **padišx^war** ‘dish, bowl’ likewise shows a two-consonant cluster if the phoneme /x^w/ is accepted (see § 30.4.4.9).

There are only a few cases with original three-consonant clusters like **nirfs-** prs. ‘wane, decrease’ or **murnjēn(īdan)** prs. ‘destroy’, where again Avestan influence is certain (cf. Av. *nərəfs-* and *mərənčaiti*), as in transcriptions of Av. words like **xurdruš** ‘with a bloody club’ (Av. *xrwidruš*), **xrafstar** ‘reptile, noxious creature’ (Av. *xrafstra-*), or (in final position) **huwaršt** ‘virtue, good deed’ (Av. *hwaršta-*).

27. The Pahlavi readings **dagr**, **sagr**, and **šagr** are due to the ManMP spellings *dgr* and *šgr*. The words may as well be transliterated <dl>, <syl>, and <šyl>, since the character for <d>, <g>, and <y> is identical.

28. Cf. Parth. *šrg*, Manichæan Sogdian *šrwγ*, Buddhist Sogdian *šrγw* ‘lion’ (Henning 1965: 45f.).

29. <c> is Iranological convention to represent the Semitic letter <ç>.

30.6. Gemination

30.6.1. Morpheme-internal

Gemination of consonants both voiced and voiceless is met with only within a word or at its end. Geminated /r/ occurs quite often, as it is the historical result of older *rn; cf. the following examples.

In medial position:

darrīdan ‘split, tear’ < *darnā-, cf. Av. ¹dar- ‘cleave’, prs. *dərənā-Farrbay* ‘the first major Fire of Sasanian Iran, that of priests and nobles’

farrox ‘fortunate, blessed, happy’ < *farnahwant-

kirrōg ‘artisan, craftsman; skilled’ < *kṛnuvaka-

parrag ‘wing’ < *parnaka-

spurrīg ‘entire, complete, perfect’ < *us-pr̥na-

tarrag ‘vegetable’ < *tarnaka-

wurrōyistan ‘believe; choose’ < *wṛ̥nau-, OInd. *vṛ̥ṇōti*

In final position:

āx^warr ‘manger, stable’ < *ā-hwar(a)na- (?)

parr ‘feather, wing’ < *parna-

warr ‘wool; garment’ < *warnā-, OInd. *ūr̥ṇā-*

zarr ‘gold’ < *zarna-

Though the geminated /r/ may have had a long pronunciation, there are nevertheless no hints of phonemic status of length, as there obviously do not exist any minimal pairs such as **warrag** : ***warag** or the like; the pair **parrag** : **pardag** shows that it is best to explain the sequences -rr-/-rd- as being biphonemic.

30.6.2. At morpheme boundaries

Obviously in all other cases where geminated consonants are involved they occur at old boundaries of different morphemes; in these cases, too, geminated consonants are the result of assimilation, e.g.

hammōxtan ‘teach; learn’, < *ham-mauk-

hannām ‘member, limb’, < *han-dāman-

hazzān ‘tomb, ossuary’ (?)

pa(č)čēn ‘copy’, < *pati-čagna-

parrēxtan ‘remain over, be superfluous’ < *pari-raik-

passāxtan ‘fashion, prepare, etc.’ < **pati-sak-*
paššinjag ‘drop’ < **pati-šinčaka-*
pattūg ‘enduring, patient, persevering’ < **pati-tav-*
pazzāftan ‘(cause to) ripen’ < **pati-ĵāmaya-*

30.6.3. By assimilation to suffixes

Gemination is also met with in words ending in a dental when suffixed by the comparative suffix **-tar** or the superlative suffix **-tom**; in such cases the final dental, which as a rule is voiced (since in this position the phonological distinction between voiceless and voiced consonants is neutralized), becomes voiceless when the final dental is etymologically voiceless:³⁰

wad ‘bad’ : **wattar** ‘worse’, **wattom** ‘worst’

This rule mostly applies also to words with final guttural³¹ to which the suffix */-gar/* < **-kara-* ‘doer, doing’ is added:³²

ardīg ‘battle’ → **ardīkkar** ‘warlike, warrior’, ManMP *ʾrdyqkr*
bazag ‘sin, evil’ → **bazakkar** ‘sinful, sinner’, ManMP *bzkr*
kirbag ‘virtue, good deed’ → **kirbakkar** ‘beneficent’, ManMP *kyrbkr*
lābag ‘supplication’ → **lābakkar** ‘suppliant’ (written <lʾpkk!>!)
zīndag ‘alive, living’ → **zīndakkar** ‘vivifier, reviver’, ManMP *zyndkr*

These are also dealt with under § 30.7.2.

In rare cases a kind of spontaneous gemination in words of diminutive character or of uncertain origin may occur:

waččag ‘child, baby, cub’ ~ NP *bačče*
wassad ‘coral’ ~ NP *bussad*

30. In the case of **kabōd** ‘gray-blue; pigeon’ : **kabōtar** ‘pigeon’, the same suffix seems to be implied (but not with comparative meaning); simple */-t-/* probably depends upon the fact that there was no semantic connection with the comparative suffix in this word any more. The long vowel before the <-t> could also have been significant, cf. the interchanging of long/short vowel plus simple/long consonant in a word like **pāsox** = **passox** ‘answer’.

31. In most instances the final guttural will have been the **-ka*-suffix.

32. These examples are well corroborated by ManMP evidence; where ManMP shows the suffix **-kara-* to be *-gr*, it must be surmised that the word to which it is added did not contain the **-ka*-suffix, cf. e.g. ManMP *qʾrgr* ‘worker’ < **kāra-kara-*, also represented by MP of the ostraca <kʾlykl> O.117b 3 (Weber 1992: 74).

30.7. Morphophonology

30.7.1. *Phonological rules ignored*

Phonological rules are often disregarded at morpheme boundaries; thus, in compound words voiceless stops (or affricates) are retained despite their voiced surroundings, cf. **a-čār** ‘inevitable’ (**čār** ‘means, remedy’), **a-pus** ‘childless’ (**pus** ‘son’), **hu-tuxš** ‘artisan’ (**tuxšīdan** ‘skill, endeavor’). In all these cases prefix and lexeme are clearly recognizable from a synchronic point of view; see also § 30.4.4.

30.7.2. *Suffixes generalized*

Certain suffixes regularly have generalized forms with voiceless consonants although, by historical development, voiced counterparts ought also to have existed. In the case of the suffixes **-tar** (comparative) and **-tom** (superlative) only these forms are in use, though ***-dar** and ***-dom** should have been the outcome of ***-atara-** and ***-atama-** originally;³³ but only in isolated words were these perhaps retained, e.g. **fradom** ‘first’ < ***fratama-**. A consequence, already discussed in § 30.6.4, of the exclusive use of **-tar** and **-tom** is the devoicing, in adjectives, of the final voiced dentals, cf. **wad** : **wattar** : **wattom**.

With regard to the suffix **-gar** (forming actor nouns) < ***-kara-**, the regular form (with initial voiced stop) has been generalized, but in cases when it is added to a word ending in **-ag** (< ***-aka-**) a geminated voiceless stop (**-kk-**) is preserved, cf. **zīndag** : **zīndakkar** (see § 30.6.3). But there are instances where this rule obviously is neglected, cf. e.g. **čārag** ‘means, remedy’ : **čārag-garīh** ‘ability’; here a voiced pronunciation seems to be indicated by <-g-> (not <-k->) in the spelling <-glyph>.

30.7.3. *Prefixes lexicalized*

Gemination often is the result of assimilation at morpheme boundaries. This applies especially to preverbs such as ***ham-** or ***pati-**. But since verbs with these prefixes, as a rule, were lexicalized in pre-Middle Iranian times the assimilation of ***ham-** or ***pati-** to the first consonant of a nominal element (cf. **hannām** < ***han-dāman-**) or a verbal stem (such as **passāxtan** < ***pati-sak-**) is not to be dealt with under morphophonology.

30.7.4. *Particles cliticized*

Morphophonological changes may be noted with certain conjunctions and prepositions when they are suffixed with enclitics, especially enclitic

33. Cf. ManMP *-dr*, but NP *-tar*.

pronouns. Thus the conjunctions **ā** ‘then’ or **ka** ‘when’ will not change their form when suffixed, cf. **ā-š** (with encl. pron. 3rd sg.) or **ka-š**. On the other hand the conjunction **ud** ‘and’ has the by-form **u-** when it is connected with enclitics: **u-m** ‘and by me’, **u-t** ‘and by thee’, **u-š** ‘and by him/her/it’, **u-mān** ‘and by us’, **u-tān** ‘and by you (pl.)’, **u-šān** ‘and by them’; this is clearly shown by the ManMP spelling ^ʔ*wd* vs. ^ʔ*wm*, ^ʔ*w_t*, ^ʔ*wš*, etc. The prepositions **az**, **ō**, and **pad** have by-forms **aziš**, **awiš**, and **padiš** which are sometimes used as postpositions with the meaning of their unsuffixed counterparts. The question whether, in these instances, the short vowel connecting the preposition and the suffixed enclitic was really [-i-] or [-e-] instead (see § 30.4.5.1) cannot be answered; in both Pahlavi and ManMP, spellings with or without <-y> are possible, which only shows that the vowel must have been short. MacKenzie prefers [-i-], but since already ENP regularly shows *-a-* (in *-aš*, *-at*, etc.) with enclitic pronouns, the readings *-eš*, *-et*, etc. for MP should have been equally possible. This holds good also for the adverbial enclitic **-iz**, written <-(y)c>, with the meaning ‘also, even’.

30.8. Stress

Nothing is actually known about stress in Pahlavi. Since the free accentuation of early Indo-Iranian must have changed greatly during the course of time in Western Iranian, because of the loss of final syllables and through the syncope of vowels in various suffixes (e.g. the *-aka*-suffix), a system of stress must have arisen that depended on the length of the word and at the same time the length of the syllables. Thus, we can only guess from various word forms that stress, presumably in course of the evolution of Middle Persian, was of great importance. From syncopated syllables it can be inferred that stress on different syllables could have different results; but the reason(s) (if there were more than one) that led to special forms in special cases are still obscure. Thus the Armenian doublet *hazarapet* and *hazaravuxt* < **hazāra-pati-* show that in the first case the lending language stressed the word as **hazā́ra-pàti-* but in the second example one has to assume stress on a different syllable, viz. **hazārá-pati-* so that the *-a-* of **-pati-* could be syncope (cf. Hübschmann 1897:174; Back 1978: 223). But as a matter of fact this development must have taken place in a period when OP was declining and changing into Middle Iranian status; so there are no indications at all that stress had a phonemic function in “classical” Pahlavi.

These considerations have led to some current models of the development of stress in Middle Persian; but since it is of no importance with re-

gard to phonology, we may refer to Back 1978, where the earlier literature is summarized.

30.9. Loanwords

MP as an official language was open to many linguistic influences both from other Iranian dialects and from surrounding non-Iranian languages. First of all, its genuine Persian (Southwest Iranian) basis was overlain by northwestern elements very early, mainly from Arsacid times, i.e. Parthian. These are clearly distinguishable by lexical, morphological, syntactic, and some striking phonological features such as the loss of initial **h-* e.g. in the preverb **ham-* (cf. **amburdan** vs. **hambārīdan** ‘fill, collect’), the keeping of initial **y-* (cf. **yāwēd(ān)** vs. **jāwēdān** ‘always, eternally’) or the different development of **-θr-* in Manichæan Parth. **šhrb** vs. **šasab** ‘satrap’ (< **xšaθrapā-*); cf. *in extenso* Tedesco 1921. It is perhaps by this impact of Parthian that some phonemes such as /y/ were innovated in the course of time, that the allophone [ž] perhaps acquired phonological status (or that /y/ was restored initially). With regard to /y/ and /ž/, one must not forget the influence Eastern Iranian languages, especially Sogdian, had on the Persian language from late Sasanian times (cf. Henning 1939). A general survey of the evolution of Middle Persian as a kind of “mixed” language is given in Henning 1958: 92f.

30.9.1. Loanwords from Avestan

Whereas there was a tendency to incorporate Parthian and other Iranian influences as completely as possible into the Middle Persian phonological system, but also by changing it, words of Avestan origin, which played an important role in the texts with their technical vocabulary of Zoroastrianism, tend to be only partially assimilated to the system. Avestan phones are represented, as far as possible, by their MP equivalents but changed if there was no such phone in MP. Thus, for example, Av. *θ* is replaced by *s* in the cluster *θr* but by *h* in intervocalic position; this may be due to parallel cases of the known development of OP **θ* > MP *h* (cf. **daθa* ‘ten’ > **dah**). Furthermore, the rules for the occurrence of phonemes within a word are often neglected (e.g. voiceless stops or affricates in intervocalic position) as well as the number of consonants possible in a cluster (cf. **Wōrūbaršt** or **xurdruš**). The following list of selected items in the Pahlavi lexicon from Avestan may illustrate the possible kinds of adaptation as found in MacKenzie 1971:

- agrē** (*ayriia-*) ‘highest, greatest’
āgrift (*āgarəpta-*) ‘crime, offense; a degree of sin’
Arzah (*arəzahī-*) ‘the eastern continent of the earth’
āsrōm (*āθrauuan-*) ‘priest’
dāsar (*dāθra-*) ‘gift, reward’
garōdmān (*garō.dəmāna-*) ‘paradise’
nask ‘a division of the Avesta’
pōryōtkēš (*paoiryō.tkaēša-*) ‘(one of) the first teachers of the Mazdean religion’
sēj (*iθyejah-*) ‘danger, trouble’
šusar (*xšudra-*) ‘liquid, fluid; semen’
***uruspar** (*uruθbar-*) ‘intestines, womb; core, essence’
us(t)ōfrīt (*ušfriti-*) ‘offering’
uzērin (*zayeirina-*) ‘the evening *gāh*’
wačast (*vačas-tašti-*) ‘strophe (of the Gathas)’
waydān (*vaydana-*) ‘head (ahuric)’
Wōrūbaršt (*vouru.barəšti-*) ‘the northwestern continent of the earth’
Wōrūjaršt (*vouru.jarəšti-*) ‘the northeastern continent of the earth’
xurdruš (*xrvidruš*) ‘with a bloody club’
X^wanirah (*x^vaniraθa-*) ‘the central, inhabited continent of the earth’
x^wēdōdah (*x^vaētvadaθa-*) ‘kin-marriage’
x^wēn-āhan (*x^vaēna-[ayah-]*) ‘shining metal’
yask ‘illness, sickness’ (or Parth.?)
yāt (*yāta-*) ‘a degree of sin’

30.9.2. *Loanwords from other languages*

Loans from other, i.e. non-Iranian, languages are rare. Though contact with the Byzantine empire was seldom on friendly terms, the Greek world gave many an expression of both worldly and philosophical substance; cf. **filāsōfā** ‘philosopher’ or **sōfistā** ‘sophist’.

Through Greek also, some other words like **yākand** ‘ruby’ (< ῥάκινθος) or names such as **Hrōm** ‘Rome’ or **Aleksandar** > ***Skandar** ‘Alexander’ (cf. NP *Iskandar*, *Sikandar*) entered the Middle Persian language; whether the name for ‘Caesar’, viz. **kēsar**, also came by way of the Greek language is doubtful as Iranians had contact with the (West) Romans from Parthian times.

From Indic there is but one word to be mentioned that penetrated Pahlavi rather early, perhaps by way of Sogdian, viz. **but** ‘Buddha’, though both

ManMP and Manichæan Parth. have far more loans from the East due to the syncretistic character of the Manichean religion where Buddhist as well as Christian elements were amalgated (cf. e.g. Sims-Williams 1983).

As a rule, only after the Islamic conquest could Arabic words have penetrated Pahlavi, e.g. **kabāh** 'garment, cloak' (NP *qabā*) < Arab. *qabā*') or **zayt** 'olive'.

Bibliography

- Abrahamjan, R. 1965. *Pexlevijsko-persidsko-armjano-russko-anglijskij slovar*. Erevan: Mitk.
- Andreas, F. C., and K. Barr. 1933. *Bruchstücke einer Pehlevi-Übersetzung der Psalmen* (Sitzungsberichte der Preußischen Akademie der Wissenschaften, Berlin, Phil.-Hist. Klasse, 1933/1).
- Back, M. 1978. *Die sassanidischen Staatsinschriften*. *Acta Iranica* 18.
- . 1981. "Die mittelpersische Lautverschiebung: Ein Stilwandel." *Die Sprache* 27/2: 178–86.
- Bartholomae, Chr. 1904. *Altiranisches Wörterbuch*. Strassbourg: Trübner (repr. Berlin: de Gruyter, 1961).
- Benveniste, Émile. 1968. "Le système phonologique de l'iranien ancien." *Bulletin de la Société de Linguistique de Paris* 63: 53–64.
- Bivar, A. D. H. 1969. *Catalogue of the Western Asiatic Seals in the British Museum: Stamp Seals*, vol. 2, *The Sassanian Dynasty*. London: British Museum.
- Bolognesi, Gi. 1960. *Le fonti dialettali degli imprestici iranici in armeno*. Milan: Vita e Pensiero.
- Boyce, M. A. 1954. *The Manichaean Hymn-Cycles in Parthian*. London: Oxford University Press.
- . 1977. *Word-list of Manichean Middle Persian and Parthian. With a Reverse Index by R. Zwanziger*. *Acta Iranica* 9a.
- Dresden, Mark J. 1970. "Middle Iranian." In *Current Trends in Linguistics*, ed. Thomas A. Sebeok, vol. 6, *Linguistics in South West Asia and North Africa*, 26–63. The Hague: Mouton.
- Édel'man, D. I. 1977. "K fonemnomu sostavy obščeiranskogo (o fonologičeskom statute *x^v)." *Voprosy Jazykoznanija*, no. 4: 81–88.
- Gignoux, Ph. 1972. *Glossaire des inscriptions pehlevies et parthes*. London: Lund Humphries.
- Hansen, O. 1963. *Mittelpersisches Lesebuch*. Berlin: de Gruyter.

- Henning, W. B. 1939. "Sogdian Loan-words in New Persian." *Bulletin of the School of Oriental and African Studies* 10: 93–106 (repr. in *Selected Papers*, Leiden: Brill, 1977, 1: 639–52).
- . 1958. "Mitteliranisch." In *Handbuch der Orientalistik*, IV/1: 20–130. Leiden: Brill.
- . 1965. "A Grain of Mustard." *Annali dell'Istituto Orientale di Napoli. Sezione Linguistica* 6: 29–47.
- Hübschmann, H. 1897. *Armenische Grammatik*, part 1, *Armenische Etymologie*. Leipzig: Breitkopf & Härtel (repr. Hildesheim: Olms, 1962).
- Kapadia, D. D. 1953. *Glossary of Pahlavi Vendidad*. Bombay: Parsi Panchayet.
- Lazard, G. 1963. *La langue des plus anciens monuments de la prose persane*. Paris: Klincksiek.
- MacKenzie, D. N. 1967. "Notes on the Transcription of Pahlavi." *Bulletin of the School of Oriental and African Studies* 30: 17–29.
- . 1969. "Iranian Languages." In *Current Trends in Linguistics*, ed. Thomas A. Sebeok, vol. 5, *Linguistics in South Asia*, 450–77. The Hague: Mouton.
- . 1971. *A Concise Pahlavi Dictionary*. London: Oxford University Press.
- . 1989. "Kerdir's Inscription." In *Naqsh-e Rostam 6, The Sasanian Rock Reliefs at Naqsh-e Rostam*, by G. Herrmann, 35–72 (*Iranische Denkmäler*, ser. 2, *Iranische Felsreliefs*, 1). Berlin: Reimer.
- . 1991. "Middle Iranian Lexicography." In *Wörterbücher/Dictionaries/Dictionnaires (Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/Encyclopédie internationale de lexicographie)*, ed. F. J. Hausmann, O. Reichmann, H. E. Wiegand, and L. Zgusta, vol. 3, pp. 2473–77. Berlin: de Gruyter.
- Martinet, André. 1965. "La linguistique synchronique." In *Études et recherches*, Paris: Presses Universitaires de France.
- de Menasce, J. 1983. "Zoroastrian Pahlavī Writings." In *The Cambridge History of Iran*, vol. 3, part 2, *The Seleucid, Parthian and Sasanian Periods*, ed. E. Yarshater, 1166–95. Cambridge: Cambridge University Press.
- Molčanova, E. K. 1978. "Sredneiranskije jazyki." In: *Jazyki Azii i Afriki. Indoevropskie jazyki. II. Iranskije jazyki, dardskie jazyki. Dravidijskie jazyki*, 58–109. Moscow: Nauka.
- Nyberg, H. S. 1928–31. *Hilfsbuch des Pehlevi*, part 1, *Texte und Index der Pehlevi-Wörter*; part 2, *Glossar*. Uppsala: Almqvist & Wixell.

- . 1964–74. *A Manual of Pahlavi*, part 1, *Texts*; part 2, *Ideograms, Glossary, Abbreviations, Index, Grammatical Survey, Corrigenda to Part 1*. Wiesbaden: Harrassowitz.
- Oranskij, I. M. 1960. *Vvedenie v iranskuju filologiju*. Moscow: Nauka.
- Pisowicz, A. 1984. "The Development of the Middle Persian System of Obstruents." In *Middle Iranian Studies: Proceedings of the International Symposium organized by the Katholieke Universiteit Leuven from the 17th to the 20th of May 1982* (*Orientalia Lovaniensia Analecta* 16), 15–24. Leuven: Peeters.
- . 1985. *Origins of the New and Middle Persian Phonological Systems*. Cracow: Nakładem Uniwersytetu Jagiellońskiego.
- Pulleyblank, E. G. 1952. "A Sogdian Colony in Inner Mongolia." *T'oung Pao* 41: 317–36.
- Rastorgueva, V. S. 1966. *Srednepersidskij jazyk*. Moscow: Nauka.
- Rastorgueva, V. S., and E. K. Molčanova. 1981. "Srednepersidskij jazyk." In *Osnovy iranskogo jazykoznanija—Sredneiranskie jazyki*, 6–146. Moscow: Nauka.
- Rossi, A. V. 1973. "On the Transcription of Pahlavi Again." *Annali dell'Istituto Orientale di Napoli* 33: 126–32.
- . 1975. *Linguistica Mediopersiana 1966–1973: Bibliografica Analitica* (*Annali dell'Istituto Orientale di Napoli*, supplement 5 to vol. 35, no. 4).
- Salemann, C. 1895. "Mittelpersisch." In *Grundriß der iranischen Philologie*, ed. W. Geiger and E. Kuhn, vol. 1, part 1, 249–332. Strassbourg: Trübner. Trans. as *A Middle-Persian Grammar*, by L. Bogdanov. Bombay, 1930.
- Schmitt, Rüdiger, ed. 1989. *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert.
- Sims-Williams, N. 1983. "Indian Elements in Parthian and Sogdian." In *Sprachen des Buddhismus in Zentralasien: Vorträge des Hamburger Symposiums vom 2. Juli bis 5. Juli 1981*, ed. K. Röhrborn and W. Veenker, 132–41. Wiesbaden: Harrassowitz.
- Skalmowski, W. 1962. "Über einige statistisch erfaßbare Züge der persischen Sprachentwicklung." *Folia Orientalia* 4: 47–80.
- Sprengling, Martin. 1953. *Third-Century Iran: Sapor and Kartir*. Chicago: Oriental Institute.
- Sundermann W. 1989a. "Westmittelliranische Sprachen." In Schmitt 1989: 106–13.

———. 1989b. "Mittelpersisch." In Schmitt 1989: 138–64.

Tafazzoli, A. 1969. *Glossary of Mēnōg ī Xrad* (Iranian Culture Foundation, 63. Pahlavi Studies, 6). Teheran: Iranian Culture Foundation.

———. 1970. "Notes pehlevies (I)." *Journal Asiatique* 87–89.

Tavadia, J. C. 1956. *Die mittelpersische Sprache und Literatur der Zarathustrier*. Leipzig: Harrassowitz.

Tedesco, P. 1921. "Dialektologie der westiranischen Turfantexte." *Le Monde Oriental* 15: 184–258.

Weber, D. 1992. *Pahlavi Papyri, Pergamente und Ostraca* (Corpus Inscriptionum Iranicarum, Textband.) London: School of Oriental and African Studies.

———. 1994. "Zur Problematik eines Phonems /x^w/ im Pahlavi." *Studia Anatolica, Mesopotamica et Iranica* (Bamberg and Prague) 1 [in press].

Widengren, Geo. 1964. "Pahlavi and New Persian: Some Interrelations." In *J. M. Unvala Memorial Volume*, 66–75. Bombay: Dorab H. Kanga.

Hindi-Urdu Phonology

Alan S. Kaye

California State University, Fullerton

31.1. On the designations Hindi-Urdu and Hindustani: Some facts on the synchronics and diachronics of the language

For purposes of this article, no distinction will be made between the Indo-Aryan language usually called Hindi-Urdu (referred to, perhaps somewhat jokingly, as Hirdu by Kelkar 1968) and some of its other perhaps less appropriate designations.¹ It is sometimes just called Hindi (earlier Hindvi, Hindwi, Hindui, or Hindawi) or Urdu (from the Turkish *ordu* ‘army, camp, army camp’ [only the first two glosses are given in Alderson and İz 1984]), but oftentimes it is also referred to as (Bazaar) Hindust(h)ani² (also spelled Hindūstānī or Hindostani). The spoken forms reflected by all four of these designations are actually rather close to one another—so close, in fact, that a native speaker of so-called Hindi proper should have little or no difficulty in communicating with a native speaker of so-called pure Urdu or Urdu proper (except for the problem of vocabulary), or either one with a speaker of what is called Hindustani (usually identified more with Urdu than Hindi). The major distinction with all of these designations has to do with script, religion, overall culture, and to a certain extent today, country, i.e., politics (see Rai 1991: 1–35 for a survey of the history).

1. AUTHOR’S NOTE: The author gratefully acknowledges the assistance of Josef Elfenbein, who made many useful suggestions on a preliminary version of this chapter. Peter T. Daniels also offered a number of improvements to it, both stylistic and substantive, for which I am thankful. The usual disclaimers apply, however.

2. Hindustani, invented by the British government in the 19th century as a quasi-administrative language, is often defined as “a combination of Hindi and Urdu.” See, e.g., Anonymous (n.d: 2). The phonetic and phonological information given in the aforementioned book and many others like it must be used with great caution. Consider (p. 10): “/q/ has a slightly softer sound than /k/, but the untrained ear cannot detect the difference. The quite correct pronunciation of /q/ can only be acquired by hearing it in Indian speech.” There is no mention of retroflexion (see § 11.4), and vowel nasalization (see § 11.9) is symbolized *ñ*/, which is erroneously defined as (p. 7): “indicates the nasal sound of N in LONG, KING, etc., but not so strong. It is, perhaps, more like the sound of N in the French words BIEN....”

Hindi speakers are usually Hindu and are currently citizens of India (or, at least, of Indian origin) and write in the Devanāgarī (or Deonagari; cf. Scholberg 1955: 2) alphabet, known for its straight-line effect or horizontal bar above many of the consonants with some vowel diacritics written above. Characters are formed which seem to hang from a horizontal line: consider क्तिाव <kitāb> ‘book’, in which ढ = <i>, क = <k>, त = <t>, ि = <a>, and व = . Only the /i/ is written as a superscript in this instance, in conjunction with the ि which is half the grapheme. This writing system is derived from Sanskrit or Old Indo-Aryan and is also used today to write other Neo-Indo-Aryan languages, such as Marathi or Nepalese. So-called pure Urdu speakers, on the other hand, would speak more or less the same language just referred to as Hindi with more or less the exact same grammar (but with many more words deriving from Persian and Arabic; see below), except that they would tend to be Muslim (and would therefore know some Persian and Koranic Arabic). They would write the language, if literate, in the Urdu script, which is basically the Perso-Arabic one with special Indo-Aryan phonological modifications, e.g., special graphemes for the phonemically distinct retroflex consonants, e.g., ڙ, ڍ for /ɽ/ and /ɖ/ respectively.

Urdu has developed with many loanwords taken from Persian (also an Indo-Iranian language) and Arabic (a Semitic language [on which see chapter 13] with a totally different structure from the Indo-Iranian languages), and it is the oldest representative of the literary form of the language called Rekhta. There are today, of course, also many loanwords from English, such as *so so* (*he*) ‘(it is) so so’, which one often hears as a response to the question *kya hal he* or *ap kaise hē* ‘How are you?’, lit. ‘How is your state?’.

An Urdu speaker need not necessarily be a Pakistani, for there are millions of Urdu speakers in India, and some refugee speakers in Bangladesh; not to mention other countries such as Afghanistan. Moreover, an Urdu speaker is probably, though not necessarily, Muslim. Hindus who speak Urdu, however, must be regarded as coming under the influence of Perso-Arabic culture.³ In fact, the center for Urdu literature and poetry is considered to be in India proper rather than in Pakistan (both Delhi and Lucknow are/were noteworthy centers in addition to the famous Aligarh).

Persian and Arabic loanwords go back to at least the times of Prithvirāja or Pithaura, who was the last ruler of Delhi before the Muslims invaded India. As a lingua franca, however, Hindi-Urdu really developed in the Delhi

3. Muhammad Iqbal, the greatest Urdu poet of them all, was originally a Kashmiri of Brahman origin.

Bazaar and spread throughout India as the Mughal Empire expanded. Naturally, during the Islamic period the number of Arabic and Persian⁴ loan-words increased significantly. Some Urdu texts can have 70% or more Perso-Arabic vocabulary, just as some Persian texts can have 70% or more Arabic vocabulary. Strangely enough, Urdu was not called Urdu until about 1800 (Masica 1991: 29).

The partition of British India into India and Pakistan in 1947 occurred by somewhat artificial and hurried means. (Here even Muhammad Iqbal—see note 3—was instrumental.) An Urdu speaker today, however, would probably come from Pakistan and align himself or herself more with Pakistani and Muslim political, socio-economic, and cultural concerns and causes, as well as Indian Muslim ones. Needless to say, this is not a hard and fast rule and there are many exceptions.

The word Urdu ['urdu:] derives from Turkish *ordu* 'army'. This fact directly points to the language's origin in the army camps during the Muslim invasions of northern India (= Hindustan) from the 8th to 13th centuries A.D. Some have derived its name from *urdū-ye muṣalla* or royal military bazaar outside Delhi Palace. Tisdall (n.d.: 1) states that Urdu "sprang into existence at Delhi through an admixture of the Persian of the Mughul conquerors with the Braj Bhāshā dialect of the conquered Hindūs. But Persian had already assimilated a great many Arabic and some Central Asiatic Turkish words. ... Persian was for some hundreds of years the court language of the Mughul Empire." The importance of Delhi in the development of Urdu is acknowledged by all investigators (Narang 1961: 13).

The vernacular around present-day New Delhi came to be known as *khārī bolī*. This variety of the language must have been some sort of *Mischsprache* between the local Indians and the foreign Arabs, Turks, Afghans, Persians, and other peoples who had come to India. This *Mischsprache* designation is even mirrored by the Hindi-Urdu designation of itself as *rekhtā*, which means 'mixed language' (also called *rekhtī*). The extent to which pidginization as a process might have influenced the development of Hindi-Urdu per se is beyond the scope of our remarks here (see Kaye 1990).

The designation 'Hindi', a Persian word deriving etymologically from the name of the river *Sindhu* 'Indus', is no doubt connected with the root for India itself (cf. *Hindūstān*, a Persian word, where the ending *-stān* means 'place of'), and Hinduism (the religion). Also, it probably was a term the invading

4. Chatterji (1960: 184) describes the Turkic conquerors (Chagatais) who mixed with Persians and invaded northern India from the 10th–13th centuries.

Muslims gave to denote the language encountered in northern India, although originally it may have referred to more than one Indo-Aryan language. Since speakers of pure Hindi tended to be associated with Hinduism as their religion and culture, this form of the language (or dialect of a general colloquial Hindustani) *tended* to borrow from Sanskrit and the later developing Middle Indo-Aryan languages rather than from the Islamic languages (Persian and Arabic). Quite naturally, borrowing from Sanskrit and Prakrit tended to influence the decision to utilize a script other than the Perso-Arabic one. Thus, for example, the pure Hindi word for 'book' is *pustāk*, which derives from Sanskrit [pustākə] (*pustaka-*), whereas Urdu used *kitab*, the word for it in Persian and Arabic as well (ultimately Arabic in origin). Hindi-Urdu tends to use the latter since it is perhaps more recognizable all over the Indian subcontinent. Similar remarks apply to Hindustani, although it is admitted that there are places in India in which *pustāk* might be more widely accepted and used rather than *kitab*. Some Hindi speakers per se might argue that the only word for book is *pustāk*, with *kitab* being pure Urdu. Urdu, however, knows only *kitab*, not *pustāk*. There is a good deal of sociolinguistic (stylistic) variation among the many varieties and styles of Hindi-Urdu. The entire matter is intricate, and our remarks here are, of necessity, an oversimplification.

There are differences between so-called pure Urdu and so-called pure Hindi in phonology, morphology, and syntax, just as there are differences between New England and southern American English (cf. Kachru 1980: 2 for distinctions among western, southern, and eastern Hindi-Urdu dialects). In phonology, one prominent distinction is that Urdu speakers have borrowed Arabic and Persian phonemes such as the voiceless uvular plosive *q* and the uvular fricatives *x* and *ɣ*, whereas Hindi uses borrowed phonemes from Sanskrit such as the retroflex *ɳ*. Then, of course, there are many lexical differences to which we have already referred.

Hindi-Urdu is spoken by over 200 million people (including those who speak it as a second language), and is thus one of the world's major languages.⁵ It has approximately 100 million speakers in Pakistan alone (the mother tongue of perhaps half of these), where it is the national language (English is the official language), and is somewhat known, chiefly by refugees, in addition to Bengali, in Bangladesh (until 1971, East Pakistan). It is also the most widely known Indo-Aryan tongue in the Dravidian south of India.

5. Chatterji (1960: 145) calls Hindusthani (his spelling) number 3 of the world's languages, preceded by number 1 Chinese and number 2 English.

Hindi-Urdu is an official language of India and has around 100 million speakers there (chiefly in northern and central India in the states of Uttar Pradesh, Madhya Pradesh, Himachal Pradesh, Bihar, Haryana, Rajasthan, and parts of the Punjab). Both Hindi and Urdu have been recognized as official languages by the Indian constitution. Further, it is widely used as a second language throughout much of India, and many use it as an auxiliary language as well. There is a Hindi-Urdu dialect known as Dakkhini (Dakhani) spoken in Hyderabad and elsewhere, which is connected with Islamic Sufism and Islamic proselytism.

Since Indians have traveled and migrated all over the world, there is a sizable population of Hindi-Urdu speakers in Trinidad, Guyana, Fiji, the West Indies, Malaysia, Indonesia, and South Africa (as is well known, Mahatma Gandhi, e.g., lived for quite some time in the last-mentioned country, as there is a large Indian emigrant colony there); however, the language is usually referred to as Hindi in these countries (for ease of identification as well as for sociolinguistic reasons). There is even a Pidgin Hindi which has developed in Fiji.

The Hindi language has a long tradition of study and numerous grammars of it were published in the 19th century. See Kachru (1980: 4–11) for a good history of such studies. The reader will note that Platts (1878), e.g., still a classic work, has been quoted in this article.

31.2. The consonants and vowels of Hindi-Urdu

The consonantal segments of Hindi-Urdu are listed in the consonant chart of Table 31-1. Retroflex is discussed in detail in § 31.4.

The vocalic segments are listed in the vowel chart of Table 31-2.

31.3. Loanword phonological characteristics

As is evident from § 31.1, Hindi-Urdu has borrowed many vocabulary items from foreign languages. The sounds [f], [z], and [s] occur only in loanwords. They are phonemic in both Persian and Arabic and are phonemic for most speakers of Hindi-Urdu as well. The phones [f] and [z] occur in Persian, Arabic, and English loanwords; e.g., *həfta* ‘week’ or *mez* ‘table’. The phone [ʃ] occurs in Sanskrit, English, Persian, and Arabic loanwords; e.g., /ʃayəd/ ‘perhaps’. In the less prestigious sociolects of the language, [ʃ] is often pronounced [s], especially by so-called Hindi speakers per se; also [f] becomes [p^h], and [z] becomes [j]. The phone [ʒ] may also occur in the speech

Table 31-1. Hindi-Urdu Consonants^a (after Ohala 1983: 2)

	Bilabial	Labiodental	Dental	Alveolar	Retroflex	Alveopalatal	Velar	Uvular	Glottal
Stops	p p ^h b b ^h		t t ^h d d ^h		ʈ ʈ ^h ɖ ɖ ^h		k k ^h g g ^h	(q)	
Affricates						ʧ ʧ ^h ʤ ʤ ^h			
Spirants		(f)		s (z)	ʃ	ʃ (ʒ)		(x) (ɣ)	h
Nasals	m		n		(ɳ)	(ɳ̌)	(ŋ)		
Laterals			l						
Flaps/Trills				r	ɽ ɽ ^h				
Approx.	w	(v) ^b				y			

a. Parenthesized items can occur in some varieties of the language, and are thus marginal phonemes.

b. [w] alternates with [v].

Table 31-2. Hindi-Urdu Vowels^a (after Ohala 1983: 1)

	Front		Central		Back	
High	i	ĩ			ũ	u
	ɪ	ĩ			ũ	u
Mid	e	ẽ	ə	ə̃	ō	o
	ɛ	ẽ			ō	o
Low	æ ^b		a	ã		

a. N.B. It is far more economical to set up a phoneme of nasalization, thus eliminating half of the inventory here. Exactly the same analysis prevails for Oriya, Assamese, Bengali, etc.

b. English loanwords only.

of educated natives who know English rather well; however, this would be rare. There are also some other marginal phones such as [ʔ] and [ɾ] which may occur in some styles and by some speakers.

31.4. Retroflexion

One of the major characteristics not only of Hindi-Urdu but also of most of Indo-Aryan (and some Iranian, e.g., Pashto), Dravidian, and Munda languages is the occurrence of retroflex consonants (often called cerebrals; cf. Scholberg 1955: 4). This is one of the major reasons why a typical Indian language of the Indo-European family sounds so very different from the European members. Chomsky and Halle (1968: 312) have termed retroflex consonants [+distributed], which are “produced with a constriction that extends for a considerable distance along the direction of air flow; nondistributed sounds are produced with a constriction that extends only for a short distance.” The voiceless retroflex stop (or plosive) *ɭ* as in Hindi-Urdu *bəɭa* ‘son’ or *bəɭi* ‘daughter’ is quite close to the allophonic pronunciation of general American English /t/ in a word like *heart*, *part*, or *port*, in which the retroflex nature of the English approximant [ɹ] = /r/ is such that the “normal” apico-alveolar /t/ retracts further back yielding a prepalatal or retroflex pronunciation.

The [tʰ] of English *top* is not a good model for the comparison to /ɭ/; the unaspirated [t] of *stop* is better. The apex of the tongue, however, is curled back (i.e., retroflexed) so that its underside touches the roof of the mouth at the mid or postpalatal area. Retroflexion is considered a place of articulation, following the IPA, though some have remarked that the “curling back” of the tongue makes retroflexion appear to be more like a manner of articulation.

Hindi-Urdu shows contrastive distribution between /t/ and /ɭ/ and all the other retroflex and non-retroflex pairs: *vo ata hɛ* ‘he comes’ vs. *vo aɭa hɛ* ‘that is whole grain flour’. Both the Devanagari writing system used to write Hindi proper and Perso-Arabic script for Urdu proper represent these different phonemes by different letters.

Two important phonological rules involving retroflex consonants are the following:

- (1) /d/ → [ɖ] / V__V
as in *bəɖa* ‘big’ or *čiriya* ‘sparrow’
- (2) /s/ → [ʂ] / __ [t] or [tʰ]
as in *ɪʂta* ‘desired’ or *čeʂta* ‘attempt’
(i.e., regressive retroflex assimilation)

Rule (1) has long been erroneously codified by the European grammarians as a free variation for all environments; i.e., *d* is “interchangeable” with *ɖ*

(cf. Platts 1878: 6). One simply has no alternative but to recognize the following four phonemes: /ɾ/, /ɾʰ/, /ɖ/, and /ɖʰ/, although one notes some free variations: *baḍi* ~ *baṛi* ‘small enclosed garden’, *uṣḍ* ~ *uṣḍṛ* ~ *uṣḍḍ* ‘fool’, or *ḍiḍkari* ~ *ḍiṛkari* ‘sudden outcry’.

Concerning rule (1), it is not correct to view [ɾ] and [ɖ] as allophones, just as it would be erroneous to claim that English [i] and [ɛ] are allophones by citing words like *economics* in which there is free variation of the initial vowel between [i] and [ɛ].

31.5. Voiced aspirated stops

Hindi-Urdu has five so-called voiced aspirated stops (i.e., murmured stops). In the literature on this topic, one usually finds some mention of the notion of breathy vowels. While the voiceless aspirated stops will not cause any problem for the English speaker, the voiced ones usually cause severe difficulties since English lacks this category. (Hindi-Urdu /h/ is voiced in many environments, i.e., [ɦ]; see rule [18].) The major problem is to avoid producing some sort of epenthetic vowel in between the voicing and the aspiration. Thus the English speaker tends to produce [bəhai] for [bʱai] or [gəhaɾa] for [gʱəɾa] ‘jar’. The voiced aspirated stops are /bʱ/, /dʱ/, /ɖʱ/, /gʱ/, and /ʃʱ/. (Kelkar 1968: 22 also lists /mʱ/, /nʱ/, and /lʱ/; see below.) They tend to show less aspiration before consonants and in final position. Thus *səməʃʱna* ‘to understand’ shows less aspiration for /ʃʱ/ than a /ʃʱ/ in initial position (e.g., /ʃʱɔpri/ ‘hut’). Similarly, the /dʱ/ of *dudʱ* ‘milk’ is weakly aspirated in comparison to the same phoneme in initial position (e.g., /dʱo/ ‘wash’).

The following minimal or near-minimal pairs illustrate the important contrastive nature of aspiration in general, a feature inherited from Proto-Indo-European (at least according to its traditional reconstruction; some linguists believe, following Hopper (1973), that *PIE had ejectives instead of voiced aspirated stops).

/pul/	‘bridge’	/sat/	‘seven’	/ɖak/	‘mail’
/pʰul/	‘flower’	/satʰ/	‘with’	/ɖʱaʃ/	‘two and a half’
/bat/	‘thing’	/do/	‘two’	/ka/	‘of’
/bʱat/	‘cooked spice’	/dʱo/	‘wash’	/kʱa/	‘eat’
		/tik/	‘teak’		
		/tʰik/	‘O.K., good’		

It seems to me that Kelkar’s (1968: 23) examples of *kumʱar* ‘potter’ (cf. Sanskrit *kambhar*), *junʱai* ‘moonlight’, and *kulʱa* ‘half of rump, buttock’ can

better be handled as a consonant cluster of the liquid + /h/ type. Most authors do not agree with Kelkar on this point and thus do not cite /m^h/, /n^h/, or /l^h/ as individual phonemes.

A few special remarks are in order concerning the intricate nature of murmured stops. Ladefoged (1971: 13) has pointed out that most Indo-Aryan languages have a contrastive series of stops with murmured release, i.e., what we have dubbed (as is customary) voiced aspirated stops. He studied the phenomenon in Hindi, Sindhi, Marathi, Bengali, Assamese, Gujarati, Bihari, Marwari, and other languages in great detail, concluding that murmured stops could also be called “breathy-voice” stops in which the arytenoid cartilages are apart while the ligamental vocal folds vibrate. A diacresis under /b/ represents a murmured /b̤/. As Ladefoged maintains (*ibid.*): “Phonemically it may be very convenient to symbolize these sounds as /b b^h p p^h/, and so on; but when one uses such a term as voiced aspirated, one is using neither the term voiced nor the term aspirated in the same way as in descriptions of the other stops.”

Ladefoged (1971: 96) also cautions us concerning the Chomsky and Halle (1968) feature “heightened subglottal pressure” for the voiced aspirated stops by noting:

Chomsky and Halle require the feature heightened subglottal pressure in their description of the murmured stops (voiced aspirated stops, in their terminology). My own experimental observations on similar languages ... cannot be interpreted unambiguously. There is sometimes, but not always, an increase in the subglottal pressure toward the end of murmured stops. This might be due to an “independently controllable aspect of the speech event” as Chomsky and Halle require for their features. But it might be simply as a result of variations in the state of the glottis, and not independently controllable in any way.

31.6. The abstractness controversy and evidence from Hindi-Urdu

Ohala (1983: chap. 3—an updated version of Ohala 1974) discusses some of the issues involved in abstract vs. concrete phonology. She offers evidence from Hindi-Urdu to help prove that there is a need for abstract underlying representations, which rarely (if ever) show up in purely phonetic form and which are psychologically real.

She explains (1983: 27) that the abstractness controversy began with Paul Kiparsky's unpublished 1968 manuscript "How Abstract is Phonology?" (first appearing in print in 1973), in which he argued that generative phonologists possessed little evidence to posit abstract underlying units. Phonologists such as Larry Hyman, Charles Kisseberth, and Michael Brame all wrote articles in favor of allowing abstract units to be postulated, citing evidence from a wide variety of languages. After these articles were published, Kiparsky rethought his position and came out in favor of abstractness in an article published in 1972 (see Ohala's references). Many years later and in retrospect, one can agree with Ohala (*ibid.*), who comments: "It seems to me that the important question is not whether the underlying forms should be abstract or concrete, but rather, what evidence there is for the 'psychological reality' of any posited underlying form whether highly abstract or not."

The material Ohala concentrates on has to do with data involving shwa-deletion; e.g., *nəmək* 'salt' but *nəmkin* 'salty'. The postulated rule is as follows:

$$\text{ə} \rightarrow \emptyset / \text{VC_CV}$$

She continues her argumentation (*ibid.*) by finding some evidence of the psychological reality of the shwa; i.e., some suffixes "seem to block the application of the shwa-deletion rule given above, e.g., [-tʲa]."

Ohala then devised an experiment involving 27 informants in which they were presented words in random order to test whether or not a shwa was added or retained. The conclusion of the experiment supports the psychological reality of abstract underlying representations. See Ohala (1983) for further details.

31.7. Hindi-Urdu and comparative Indo-Aryan phonology

Pattanayak (1966) lists cognates between Oriya, Assamese, Bengali, and Hindi in order to reconstruct their protolanguage via a strict comparative method. The author has called this protolanguage Proto-Colloquial OABH, which does not look at all like Sanskrit. Although 545 cognate sets are discussed, I present only 12 of these in Table 31-3 so that the reader may discover the similarities and differences.

Although one may easily see the archaic nature of Hindi as opposed to the other three languages compared, there appear to be some anomalies. For example, the Hindi long vowel in 'flower' is secondary and derives from a proto short vowel; however, the Hindi long vowel in 'grass' is original. Hindi consonantal gemination in 'leaf' is original but secondary in 'dog'.

Table 31-3. Comparative Indo-Aryan Phonology (Pattanayak 1966: 27)

Gloss	Oriya	Assamese	Bengali	Hindi	Sanskrit	Reference ^a
bark (tree)	c ^h eli	sal	c ^h al	c ^h a:l	chálī	T5005
flower	p ^h ulo	p ^h ul	p ^h ul	p ^h u:l	phúlla	T9092
grass	g ^h aso	g ^h āh	ghaṣ	g ^h a:s	ghāsá	T4471
leaf	pōtro	pat	pata	patta:	páttra	T7733
fruit	p ^h ōḷo	p ^h ōl	p ^h ōl	p ^h al	phála	T9051
branch	ḍaḷo	dal	ḍal	ḍa:l	ḍāla	T5546
bird	pōk ^h i	pōk ^h i	pak ^h i	pakṣi ^b	pakṣin	T7636
dog	kukuro	kukur	kukur	kutta:	kurkura	T3329
fish	mac ^h o	mas	mač ^h	mač ^h li:	mátsya	T9758
snake	sapo	xap	šap	sarp-sā:p	sarpa	T13271
elephant	hati	h(ā)ti	hati	ha:t ^h i:	hastin	T14039
horse	g ^h oḍa	g ^h ora	g ^h oḍa	g ^h oṛa:	ghoṭa	T4516

a. The Sanskrit cognates are all taken from Turner (1962ff.) courtesy of Josef Elfenbein.

b. There is a variant *pakhi*.

31.8. Gemination

Gemination or consonantal lengthening, which is crucial in Arabic (less so in Persian, but still distinctive and an important feature in the higher registers of the language, as in *bæččé* ‘child’, and of course in its many Arabic loanwords), is a phonemic feature of Hindi-Urdu (cf. for a parallel some formal English pronunciations of *bookcase*, *unnecessary*, or *unknown*, since degemination usually applies in these and other parallel cases, such as *immature* [Harris 1994: 20]). One must, therefore, either double the phonemic inventory (with a few exceptions) or consider gemination as a long component, viz., /C:/ . Although I prefer the latter solution, geminates are here written CC rather than C: or C^h.

Consider the following examples (many adapted from Bailey 1956: xix, but see § 31.11 for further remarks).

<i>pəta</i>	‘address’	<i>pətta</i>	‘leaf’	<i>məkk^hən</i>	‘butter’
<i>billi</i>	‘cat’	<i>bəggi</i>	‘buggy’	<i>šəffaf</i>	‘dear’
<i>bəttis</i>	‘thirty-two’	<i>dɪbba</i>	‘box’	<i>mʊəyyən</i>	‘fixed’
<i>kutta</i>	‘dog’	<i>bunna</i>	‘weave’	<i>həḍḍi</i>	‘bone’

<i>təʔtu</i>	'pony'	<i>təʔta</i>	'screen'	<i>gəʔdi</i>	'bundle'
<i>zimmə</i>	'responsibility'	<i>həmmam</i>	'bath'	<i>(h)əqda</i>	'central place'
<i>nəkku</i>	'wicked'	<i>əssi</i>	'eighty'	<i>gudda</i>	'doll'
<i>pəlli</i>	'village'	<i>hıssa</i>	'share'	<i>zıddi</i>	'stubborn'
<i>bədd^ha</i>	'bound'	<i>mʊəzzin</i>	'muezzin'	<i>nəddi</i>	'river'

In these data, the following rules are significant:

- (3) /t/, /h/, /b^h/, /s/, /f/, and /y/ cannot be geminated.
- (4) Although the orthography distinguishes geminated consonants in final position (e.g., *mlečč^h* 'barbarian' or *nirləjj* 'shameless'), there is a degemination present phonetically. Perso-Arabic loanwords are also degeminated in final position; e.g., *haq* ~ *hak* 'true'. This probably represents a colloquial Arabic feature, since in many colloquial Arabic dialects there is a neutralization of gemination in final position. This is not a characteristic of Classical Arabic, in which this particular word is pronounced *haqq*. One may also cite Hindi-Urdu *rəbb* 'God' as another example of this, viz., /rəb/. Cf. Classical Arabic /rabb/.
- (5) There are some free variants involving geminates; e.g., 'Delhi' can be *dilli* or *dəhli*; the interjection 'Oh!' is *ohho* ~ *oho*. See also § 31.11.
- (6) In the case of geminated aspirated stops, the aspiration occurs at the end of the geminated consonant and is transcribed thus here; i.e., the raised ^h occurs only after the second member.
- (7) Geminated /w/ → /vv/ (e.g., *əvvəl* 'first'), although not for all speakers.

31.9. Nasalization

Oral and nasalized vowels contrast in such minimal pairs as *ak* 'plant' vs. *āk* 'draw', *sas* 'mother-in-law' vs. *sās* 'breath', or *bas* 'bad smell' vs. *bās* 'bamboo'. Although some authors have argued against the phonemic status of [ē] and [ō], claiming that these are, in essence, allophones of [ē] and [ō], respectively, the following minimal set contradicts this: [mē] 'in' vs. [mē̃] 'I' vs. [mɛ̃] 'wine'.

Nasalized vowels frequently distinguish singular verbs from plural ones. Consider: /hɛ/ 'is' vs. /hē/ 'are' or /čəli/ 'she went' vs. /čəlī/ 'they (f.) went'. Vowel nasalization can also mark the indicative vs. subjunctive; e.g., /čəle/ 'they (m.) went' vs. /čəlē̃/ 'shall we go (subj.)'.

There are words which display free variation between nasalized and non-nasalized vowels, e.g., /dəs/ ~ /dəs/ 'bite'. There are also free variants of /VC/ ~ /ṼnC/, e.g., /həsi/ ~ /hē̃si/ 'laughter'.

The following rules are operative for many varieties of Hindi-Urdu:

- (8) A long nasalized vowel may be followed only by a voiceless stop and not a voiced one.
- (9) The sequence V + [+stop, +homorganic] + [+stop, -voiced] is a mark of a loanword.
- (10) Although there are exceptions, / $\tilde{V}$ / $\rightarrow$ [-nasalized] / [+obstruent]__.
- (11) As in English, there is an allophonic rule nasalizing vowels before nasal consonants; e.g., [nām] 'name'. However, this may also occur, in contradistinction to English, after nasal consonants as well; e.g., [nāk] 'nose'.
- (12) The sequence V + [+nasal, +voiced, +homorganic] represents more nasalization than the same vowel before [-nasal, -voiced, +stop].
- (13) Nasals tend to be homorganic before stops; e.g., *tamba* 'copper', *gend* 'ball', *p^həŋki* 'handful', and *j^həŋḍa* 'flag'. Exceptions may seem to occur due to schwa-deletion of an underlying schwa.

31.10. Some further allophonic rules

The following rules apply to the phonemes of Hindi-Urdu and establish various allophones:

- (14) /n/ and /l/ $\rightarrow$ [+alveolar] / [+retroflex]
- (15) /r/ $\rightarrow$ [+fronted] / [+dental]__
- (16) [+stop, +voiced] $\rightarrow$ [+prenasalized] / / $\tilde{V}$ /__
- (17) /r/ $\rightarrow$ [+lateralized] / __[l]
/r/ $\rightarrow$ [+fricativized] / [+consonantal, -voiced]__
- (18) /h/ $\rightarrow$ [h] / __# (e.g., [rah] 'path')
/h/ $\rightarrow$ [ɦ] elsewhere
Bender (1967: 26) remarks that /h/ is always voiced ("never voiceless as in English"). Ohala (1983: 7) states that medial /h/ can be voiced or voiceless.
- (19) [+vocalic] $\rightarrow$ [+retroflex] / [+retroflex]
- (20) [+vocalic] $\rightarrow$ [+higher] / /CV/
[+vocalic] $\rightarrow$ [+lower] / /VC/
- (21) / $\tilde{V}$ / $\rightarrow$ [ṽ] / __v/ (optional rule)

31.11. Consonant cluster simplification and compensatory vowel lengthening

There are quite a few cases in which geminated consonants have free variation forms with non-geminates; but where this does occur, it often requires a compensatory lengthening of the preceding vowel: *mak^hən* ~ *makk^hən* ‘butter’, or *gid^h* ~ *gidd^h* ‘vulture’. Evidence such as this can be successfully used to demonstrate the structural affinity between /a/ and /ə/, /i/ and /ɪ/, which can be written as /ā/, /a/, /ī/, and /i/ respectively.

Some authors have claimed that aspirated stops are not geminated; e.g., Gumperz and Naim (1960: 103). However, the evidence does not favor their statement. There are also diachronic doublets to which the above should be compared: *bur^ha* ~ *budḍ^ha* ‘old man’, *b^hāg* ~ *b^hang* ‘bhang; hemp’. Cases of simple free variation between geminates and non-geminates occur as well: *ṣākər* ~ *ṣakkər* ‘sugar’, or *uṭ^ha* ~ *uṭṭ^ha* ‘arose’.

31.12. Stress

Stress is not phonemic, and there is a considerable degree of free variation. Most words, however, take penultimate stress inherited from Middle Indo-Aryan. Old Indo-Aryan had pitch accent which then developed into stress. The details are quite complicated. Scholberg (1955: 14) remarks that the more monotonic the Hindi-Urdu is, the better. Thus a sentence like *mē kam kərta hū* ‘I am working’ should be pronounced ~~~~~, or *vo pani pita he* ‘he is drinking water’ should be ~~~~~ (ibid.).

31.13. Consonant cluster constraints

Appendix 3 of Ohala (1983: 168–83) presents examples of consonant clusters in various environments. Triple consonant clusters are about as complicated as the situation gets in initial position. These clusters must begin with /s/ followed by a stop. In medial position there are also numerous examples of triple consonant clusters. So too for final position, although there is one example cited for a quadriconsonantal cluster, viz., *wəṛtsy* ‘alveolar’ (a Sanskrit loanword).

31.14. Vowel cluster constraints

Ohala (1983: 72ff.) did not study this topic with the same vigor as the topic of the previous section. Very few diphthongs are observable in initial posi-

tion: [ai], [au], and [ua]. In noninitial position, many more vowels occur: [əi], [æe], [ui], [ai], [oi], [ei], [au], [əu], [ua], [oa], [uə], [æe], [ao], and [au]. [au] and [uə] do not occur finally.

31.15. The structure of monosyllabic morphemes

Ohala (1983: 55) has demonstrated that monosyllabic native syllables are $C_0^2VC_0^2$. Examples are as follows:

V	[a]	'come'	CVCC	[bʰaŋg]	'bhang, hemp'
CVC	[kəl]	'tomorrow, yesterday'	CCVC	[pyas]	'thirst'
VC	[ag]	'fire'	CCV	[kyō]	'why'
CV	[lō]	'flame'	VCC	[əst]	'set'

However, if loanwords are considered, the structure then becomes $C_0^3VC_0^3$.

References

- Alderson, A. D., and Fahir İz. 1984. *The Oxford Turkish–English Dictionary*, 3rd ed. Oxford: Oxford University Press.
- Anonymous. n.d. *Hindustani Simplified*. London: Hugo Language Institute.
- Bailey, T. Grahame. 1956. *Teach Yourself Urdu*. London: English Universities Press. [Revised edition of J. R. Firth and A. H. Harley, *Teach Yourself Hindustani*. London: English Universities Press, 1950.]
- Bender, Ernest. 1967. *Hindi Grammar and Reader*. Philadelphia: University of Pennsylvania Press.
- Chatterji, Suniti Kumar. 1960. *Indo-Aryan and Hindi*. Calcutta: Firma K. L. Mukhopadhyay.
- Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.
- Gumperz, John J., and C. M. Naim. 1960. "Formal and Informal Standards in the Hindi Regional Language Area." In *Linguistic Diversity in South Asia*, ed. C. A. Ferguson and J. J. Gumperz, pp. 97–118. Bloomington: Indiana University Press.
- Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell.
- Hopper, Paul J. 1973. "Glottalized and Murmured Occlusives in Indo-European." *Glossa* 7: 141–66.
- Kachru, Yamuna. 1980. *Aspects of Hindi Grammar*. New Delhi: Manohar.

- Kaye, Alan S. 1990. "Observations on Pidginistics and Creolistics." *Semiotica* 78: 285–348.
- Kelkar, Ashok R. 1968. *Studies in Hindi-Urdu I: Introduction and Word Phonology* (Deccan College Building Centenary and Silver Jubilee Series, 35). Poona: Deccan College, Postgraduate and Research Institute.
- Kiparsky, Paul. 1968. "How Abstract Is Phonology?" ms. In *Three Dimensions of Linguistic Theory*, ed. O. Fujimura, pp. 5–56. Tokyo: TEC, 1973.
- Ladefoged, Peter. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.
- Masica, Colin P. 1991. *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.
- Narang, Gopi Chand. 1961. *Karkhandāri Dialect of Delhi Urdu*. New Delhi: Caxton Press.
- Ohala, Manjari. 1974. "The Abstractness Controversy: Experimental Input from Hindi." *Language* 50: 225–35.
- . 1983. *Aspects of Hindi Phonology* (HLBD Series in Linguistics, 2). Delhi: Motilal Banarsidass.
- Pattanayak, Debi Prasanna. 1966. *A Controlled Historical Reconstruction of Oriya, Assamese, Bengali and Hindi*. The Hague: Mouton.
- Platts, John T. 1878. *A Grammar of the Hindustani or Urdu Language*. Delhi: Munshiram Manoharlal.
- Rai, Amrit. 1991. *A House Divided: The Origin and Development of Hindi-Urdu*. Delhi: Oxford University Press.
- Scholberg, H. C. 1955. *Concise Grammar of the Hindi Language*. London: Oxford University Press.
- Tisdall, W. St. Clair. n.d. *A Conversation-Grammar of the Hindustani Language*. New York: Ungar.
- Turner, R. L. 1962–. *A Comparative Dictionary of the Indo-Aryan Languages*. London: Oxford University Press.

Gujarati Phonology

P. J. Mistry

California State University, Fresno

My purpose is threefold: (i) to present some salient characteristics of the Gujarati sound system and its relationship with other levels of the language, (ii) to report previous studies and discuss issues that they raise, and (iii) to suggest an alternative analysis that resolves some of the problems of earlier accounts and reveals patterns and regularities of the language. The study additionally strives to add data from southern Gujarati (Guj.), which previous investigators have not examined.

The paper begins with a general and historical setting of Guj. § 32.2 deals with an inventory of consonants and vowels with brief discussion concerning their articulation, distribution, and realization. § 32.3 describes processes that exemplify the connection between phonology and other levels of the language. The discussion of clustering and compounding and their role in differentiating Sanskritic vs. non-Sanskritic Guj. in § 32.4 is followed by an appraisal and reanalysis of murmur, aspiration, and gemination in § 32.5. § 32.6 extends the study to the southern dialect and § 32.7 contains some concluding remarks.

32.1. Gujarati: General remarks

Guj. is an Indo-Aryan language with more than forty million speakers. It is one of the 16 languages recognized in the Indian constitution and is a *lingua franca* for the state of Gujarat. It is located on the western coast of India with Marathi to the south and Rajasthani and Marwari to the north. Except for medicine and law, it is used as a medium of instruction for all subjects

AUTHOR'S NOTE: I would like to thank Alan Kaye for his extremely valuable comments on an earlier version of this study and Fred Brengelman and Joel Nevis for their assistance in preparing the final draft of this paper.

The following abbreviations are used: caus., causative; fem., feminine; fut., future; Guj., Gujarati; IA, Indo-Aryan; impf., imperfect; inf., infinitive; masc., masculine; neu., neuter; pass., passive; pf., perfective; pl., plural; plpf., pluperfect; pres., present; Skt., Sanskrit; sg., singular; I, first person; II, second person; III, third person.

and at all levels. Its literary history extends more than a millennium and contains versified and prose writings. Voluminous publications appear in the form of newspapers, popular magazines, books, and journals on literary, critical, and expository topics. It is a language of radio and TV programs and of plays and movies.

Guj. speakers have been in the forefront in commercial and trade activities both within and outside India. Consequently, significant numbers of Gujaratis have settled in major towns and cities throughout India and in many other parts of the world, where they display their language loyalty through language instruction and literary publications. Scores of Guj. Literary Associations exist in cities outside India such as New York, London, and Karachi.

As “the language of Gandhi” (Cardona 1965: 5), one finds not only a substantial number of Mahatma Gandhi’s writings in the language but also an acceptance of a “Gandhian style,” characterized by simple sentence constructions and a minimum number of Sanskrit loans. Gandhi also introduced the current spelling convention in 1936 at the twelfth meeting of the Guj. Sahitya Parishad (Gujarati Literary Society).

Guj. is a derivative of Sanskrit (Skt.) and thus is a sister of Bengali, Hindi, Nepali, Sinhalese, and other IA languages. There are differing accounts of the historical development of the IA languages. The traditional practice is to divide their history into three stages: Old IA (Vedic and Classical Skt.), Middle IA (various Prakrits and Apabhramshas) and New IA (modern languages such as Bengali, Hindi, Kashmiri, etc.). Sir George Grierson on the other hand, attributing some of the differences in the modern IA languages to an earlier split, divides them into “inner” (North-Central languages) and “outer” (other) groups. Others (Chatterji 1926, Bloch 1965), relying on the differences in Ashokan inscriptions and other older documents, have posited a division into Eastern versus Western.

Continuous borrowing from neighboring languages, population movement, and shifting centers of the prestige language make successive marking off of descendent languages extremely difficult. Certain lines of development are, however, conjectured on the basis of innovations in phonology and grammar and their chronology. Dave (1948: 25–28) and Pandit (1973: 51–60) present the emergence of the modern languages in terms of successive family tree splits. The earliest differentiation in IA is assumed to be into Northern, Eastern, and Central varieties on the basis of such innovations as stops becoming voiced in the Northern (Sanskrit *dāntā* ‘tooth’ > Panjabi *dand* [Guj. *dāt*]) and dental and retroflex sibilants merging with the palatal

in the Eastern (Sanskrit *sandhya* ‘evening’ > Bengali *śaṭṭhā*, [Guj. *sāṭṭhā*]). Gujarati-Rajasthani, through its innovation of auxiliary verbs and postpositions, split from the Central in the eleventh century. The development of such characteristics as auxiliary *ch-* and the possessive marker *-n-* during the fifteenth century made Guj. distinct from Rajasthani.

The small sampling of data given below from three neighboring languages exemplifies the complex lines of development shaping these languages.

(1) **Past Tense**

Marathi	<i>Raj a-el</i>	‘Raj came’
Hindi	<i>Raj a-ya</i>	
Gujarati	<i>Raj av-yo/av-elo</i>	

Marathi with the *-l-* past, and Hindi with the *-y-* past correspond to Grierson’s “outer” and “inner” groups. For Guj., the *-y-* past is attributed to its membership in the “inner” group and the *-l-* past to the influence of the neighboring language, Marathi. The examination of future-forms in these languages presents a different picture.

(2) Marathi	<i>Raj ja-il</i>	‘Raj will go’
Hindi	<i>Raj ja-ega</i>	
Gujarati	<i>Raj jə-ṣe</i>	

The *-l-* future in Marathi represents a morphological innovation, the *-ega* future in Hindi an innovation of a paraphrastic construction, and the *-ṣ-* future in Guj. a retention of an element that existed in Old IA and possibly even earlier.

The prestige associated with Skt. and contact with speakers of Arabic, Persian, Portuguese, and English have influenced Guj., specifically its vocabulary. It also has words of unknown origin suspected to be survivals from indigenous languages existing at the time of the Aryans’ arrival in the region.

32.2. Consonants and vowels

Guj. consonants have six contrasting places of articulation, exhibited by the following.

(3)	bilabial	<i>bər</i>	‘successful’	palatal	<i>jər</i>	‘wealth’
	dental	<i>dər</i>	‘rate’	velar	<i>gər</i>	‘kernel’
	retroflex	<i>ḍər</i>	‘fear’	glottal	<i>hər</i>	‘each’

Table 32-1 shows the consonant phonemes.

Table 32-1. Gujarati Consonants

	Lab.	Dent. ^a	Retr.	Pal.	Vel.	Glott.
Stop [-vd, -asp]	p	t	ʈ	c	k	
[-vd, +asp]	p ^h	t ^h	ʈ ^h	c ^h	k ^h	
[+vd, -asp]	b	d	ɖ	j	g	
[+vd, +asp]	b ^h	d ^h	ɖ ^h	j ^h	g ^h	
Fricative		s		ʃ		h
Nasal	m	n	ɳ	(ɳ̃)	(ŋ)	
Lateral		l	ɭ			
Flap		r				
Glide	v			y		

a. Among the sounds listed under “dental,” only stops are dental in articulation, but *n*, *l*, *s*, and *r* are alveolars. This distinction accounts for the variance in Guj. speakers’ pronunciation of English alveolars. For instance, in *ten* ‘ten’, *ɖile* ‘delay’, and *ʈənəl* ‘tunnel’ retroflexes replace only stops but not *n* or *l*. Similar differentiation for stops and non-stops exists in the Guj. spoken among the Parsis (an Iranian community in the region for more than a millennium) who pronounce *manek əne moṭi* for standard Guj. *maṇek əne moṭi* ‘a ruby and a pearl’.

The vowels show a three-way distinction in height and place.

- (4) *mil* ‘textile factory’ *curi* ‘crushed’
mel ‘put down’ *məlməl* ‘muslin’ *cori* ‘theft’
cəri ‘grazed’
məl ‘dirt’ *cəri* ‘alter’
mal ‘luggage’
cari ‘was grazed’

They additionally contrast in terms of nasalization and murmur.

- (5) [-nasal, -murmur] *mər* ‘blossom’
[-nasal, +murmur] *mɔr* ‘gold coin’
[+nasal, -murmur] *mɔ̃g^{hi}* ‘expensive’
pɔ̃k ‘parched grain’
[+nasal, +murmur] *mɔ̃* ‘face’
rɔ̃ci ‘stupid’

The vowel phonemes are shown in Table 32-2.

Table 32-2. Gujarati Vowels

i		u
e	ə	o
(ɛ)		(ɔ)
a		

32.2.1. Stops

Stops are differentiated by the presence or absence of voicing and aspiration:

- | | | | | | |
|-----|-------------|-------------|--------------|-------------|---------------|
| (6) | [-vd, -asp] | <i>kəṇ</i> | ‘particle’ | <i>tan</i> | ‘high spirit’ |
| | [-vd, +asp] | <i>kʰəṇ</i> | ‘to scratch’ | <i>tʰan</i> | ‘breast’ |
| | [+vd, -asp] | <i>gəṇ</i> | ‘to count’ | <i>dan</i> | ‘giving’ |
| | [+vd, +asp] | <i>gʰəṇ</i> | ‘hammer’ | <i>dʰan</i> | ‘food’ |

[+voice, +aspiration] shows variation, as illustrated in the following.

- | | | | | |
|-----|---------------------------|---------------|---------------------------|----------|
| (7) | <i>dagʰ</i> ~ <i>dahg</i> | ‘blot’ | <i>labʰ</i> ~ <i>lahb</i> | ‘gain’ |
| | <i>dajʰ</i> ~ <i>dahj</i> | ‘sympathy’ | <i>badʰ</i> ~ <i>bahd</i> | ‘hurdle’ |
| | <i>daḍʰ</i> ~ <i>dahḍ</i> | ‘molar tooth’ | | |

Such variation, namely $VC^h > VhC$, is not shown by [-voice, +aspiration].

- | | | | | |
|-----|--------------|--------------------|-------------|---------------|
| (8) | <i>lakʰ</i> | ‘hundred thousand’ | <i>hatʰ</i> | ‘hand’ |
| | <i>pacʰo</i> | ‘again’ | <i>mapʰ</i> | ‘forgiveness’ |
| | <i>ʈʰaʈʰ</i> | ‘pomposity’ | | |

The variation is also blocked if the preceding vowel is nasalized.

- | | | | | | | |
|-----|-------------|-----------|-------------|--------|-------------|-------|
| (9) | <i>sājʰ</i> | ‘evening’ | <i>sāḍʰ</i> | ‘bull’ | <i>bāḍʰ</i> | ‘tie’ |
|-----|-------------|-----------|-------------|--------|-------------|-------|

The palatal stops have a spirant release and thus are really affricates. They are nevertheless represented as single segments for pattern congruity. Further, Cardona (1965: 23) maintains that bilabial, dental, and velar voiced stops spirantize intervocalically when preceded by a murmured vowel. Vyas (1977: 31–50) reports spirantization of voiceless aspirates (except the palatal one) intervocalically.

32.2.2. Sibilants

The distribution and realization of the sibilants bring to light a three-way division in the speech community. Standard Guj. has two sibilants: *s* and *š*.

ʃ occurs prominently in loanwords. Contrasting pairs for *ʃ* and *s* (*ʃur* ‘brave’ and *sur* ‘tone’) are very few.

An expanded system exists among some speakers, who have a voiced sibilant *z* in words such as *nizam* ‘administrator’, *azadi* ‘freedom’, *viziṭ* ‘visit’, and *zərina* ‘a girl’s name’, which are either names or loanwords from Arabic, Persian, and English. In these words, *z* may be replaced by *j* by some speakers. As Cardona (1965: 25) points out, it is possible to distinguish two types of speakers: the conservatives, who retain *z*, and innovators, who use *j*. An *ʃ* occurs as a first member of a cluster with *t* in loans from Skt. (*viṣiṣṭə* ‘unique’) and English (*liṣṭ* ‘list’).

In some dialects, *ʃ* is replaced by *s*, resulting in a single sibilant. Sibilants in these dialects as well as those retaining *ʃ* are optionally replaced by *h*.¹

(10)	Standard	Dialectal	
	<i>ʃū</i>	<i>hū</i>	‘what’
	<i>sat</i>	<i>hat</i>	‘seven’
	<i>dəs</i>	<i>dəh</i>	‘ten’
	<i>kəʃū</i>	<i>kərhū</i>	‘we will do’

This replacement is constrained by linguistic factors and is additionally restricted to certain registers.² The usage of these three variants in some dialects can be generalized in this way: *ʃ* by educated groups, *h* by uneducated groups, and *s* by semi-educated groups as in (11).

(11)	Educated	Semi-educated	Uneducated	
	<i>ʃak</i>	<i>sak</i>	<i>hak</i>	‘cooked vegetables’
	<i>ʃakər</i>	<i>sakər</i>	<i>hakər</i>	‘sugar’

32.2.3. Nasals

Descriptions of Guj. (Pandit 1957, Cardona 1965, Vyas 1974) report the following nasals.

(12)	bilabial	<i>dʱum</i>	‘much’	<i>rəmyə</i>	‘beautiful’
	dental	<i>dʱun</i>	‘tune’	<i>kənya</i>	‘bride’
	retroflex	<i>dʱuŋ</i>	‘to shake’	<i>əgəɳyə</i>	‘many’

1. Cardona (1965: 25) describes this *h* as a laryngeal spirant. It does not become a murmur and thus differs from the *h* discussed in § 32.5.1.

2. What is meant by “linguistic factor” is that sibilants in recent loans are retained. Some words are pronounced with sibilants at all levels and by all groups. As far as I know, there is no dialect completely devoid of sibilants. The exact conditioning factor of the deoralization of sibilants is not clear at present.

Retroflex *ŋ* has two allophones: flapped and unflapped. Unflapped *ŋ* occurs before retroflex stops, and flapped *ŋ* occurs intervocalically. In final position, flapped and unflapped *ŋ* vary freely.

A fourth nasal *N* is postulated for : (i) *ŋ* occurring between *ə* and a velar stop: *pəŋkʰo* ‘a fan’ and *əŋguli* ‘finger’; (ii) *ñ* occurring between *ə* and a palatal stop: *səñco* ‘machine’ and *əñjəli* ‘tribute’; and (iii) [*~*] occurring with other vowels: *ākʰ* ‘eye’ and *tyā* ‘there’. In the last case the nasalized vowel optionally is followed by a homorganic nasal with the following palatal or velar stop: *hīcko* ~ *hīñcko* ‘a swing’.

32.2.4. *y, v, and h*

The glides *y* and *v* are the consonantal counterparts of the two high vowels *i* and *u* and are found in initial, medial, and final positions.

- | | | | | |
|------|-------------------------|------------|-------------------------|------------------|
| (13) | <i>yad</i> | ‘memory’ | <i>vad</i> | ‘-ism’ |
| | <i>pəysa</i> | ‘money’ | <i>kevdə</i> | ‘type of flower’ |
| | <i>koi</i> ~ <i>koy</i> | ‘somebody’ | <i>beu</i> ~ <i>bev</i> | ‘both’ |

y and *v* occur in word-initial position. The sequences **vu*, **vo*, **yi*, **ye* are not found in word-initial position (except in Sanskrit borrowings).

The phoneme *v* has the variants *w* and *ʋ*: *v* is found word-initially and *w* elsewhere. In intervocalic position, *w* and *v* vary freely.

- | | | | | |
|------|-------------|-----------------------------|--------------|---------|
| (14) | <i>vrət</i> | ‘a vow’ | <i>vat</i> | ‘story’ |
| | <i>naw</i> | ‘boat’ | <i>dʰwəj</i> | ‘flag’ |
| | | <i>səwar</i> ~ <i>səvar</i> | ‘morning’ | |

h is found word-initially before a vowel, intervocalically, and word-finally.

- | | | | | |
|------|--------------|----------|--------------|--------------|
| (15) | <i>hani</i> | ‘damage’ | <i>sahəs</i> | ‘risk’ |
| | <i>kəhyū</i> | ‘said’ | <i>suleh</i> | ‘compromise’ |

(The relationship between *h* and murmur is discussed in § 32.5.1.)

32.2.5. *l, ɭ, and r*

There are two laterals: *l* and *ɭ*. The flapped *ɭ* never occurs word-initially or as a second member of a cluster.

- | | | | | |
|------|-------------|------------|-------------|-------------------|
| (16) | <i>pol</i> | ‘weakness’ | <i>poɭ</i> | ‘street’ |
| | <i>pulo</i> | ‘bridges’ | <i>puɭo</i> | ‘bundle of grass’ |

r is a flap and occurs in words like *raji* ‘pleased’, *tras* ‘oppression’, *bari* ‘window’, and *pur* ‘flood’.

32.2.6. General variation

Consonants are unreleased when they occur as the first member of a cluster: *əkʃər* ‘syllable’, and vary between released and unreleased articulation in final positions: *kədʰ* ~ *kəd* ‘size’. Single consonants contrast with geminates (exemplified and discussed in § 32.5.2 and § 32.3.3).

Vowels show qualitative and quantitative variants (Cardona 1965: 20–22).

a) They are slightly lower in an open syllable.

<i>kʊvo</i>	‘a well’	<i>kutro</i>	‘a dog’
-------------	----------	--------------	---------

b) They are lengthened in final syllable.

<i>pədoːʃ</i>	‘neighborhood’	<i>pədoːʃiː</i>	‘neighbor’
---------------	----------------	-----------------	------------

c) Nasalized vowels are lengthened.

<i>kaji</i>	‘judge’	<i>kāːji</i>	‘porridge’
-------------	---------	--------------	------------

ɛ and *ɔ* have a restricted distribution and low frequency of occurrence.

32.2.7. Stress

The stress is on the first syllable except when the second syllable has *a* and the first syllable has a vowel other than *a*: *sābu* ‘soap’, *préclit* ‘prevalent’, *tárik^h* ‘day, date’, *utáru* ‘passenger’.

32.3. Interconnections

As an exemplification of interconnections, this section sketches the phonological processes at work in the combination of morphemes (§ 32.3.1.1, § 32.3.1.2), in agreement phenomena (§ 32.3.2), and in intensification (§ 32.3.3).

32.3.1. Combination of morphemes

32.3.1.1. ə-Deletion

One widespread phenomenon prevalent in the IA languages is the deletion of *ə* before a suffix beginning with a vowel, as displayed in the following forms of the verb *cəmæk* ‘to shine’ in three languages.

(17)	Gujarati	Pres. <i>cəmək-to</i>	> <i>cəməkto</i>
		Fut. <i>cəmək-iš</i>	> <i>cəmkiš</i>
	Hindi	Pres. <i>cəmək-ta</i>	> <i>cəməkta</i>
		Fut. <i>cəmək-ūga</i>	> <i>cəm kūga</i>
	Marathi	Pres. <i>cəmək-ta</i>	> <i>cəməkta</i>
		Fut. <i>cəmək-in</i>	> <i>cəmkin</i>

The effect of the ə-deletion is additionally exhibited in the varying forms of two Guj. verbs.

(18)		‘to train’	‘to change’	
1. Inf.	<i>pəloṭ-vū</i>	<i>pələṭ-vū</i>		
2. Fut. III	<i>pəloṭ-še</i>	<i>pələṭ-še</i>		
3. Impf. masc. sg.	<i>pəloṭ-to</i>	<i>pələṭ-to</i>		
4. Pf. masc. sg.	<i>pəloṭ-yo</i>	<i>pələṭ-yo</i>		
5. Nominal	<i>pəloṭ-nar</i>	<i>pələṭ-nar</i>		
6. Imperative	<i>pəloṭ-o</i>	<i>pələṭ-o</i>	> <i>pəlto</i>	
7. Fut. I sg.	<i>pəloṭ-iš</i>	<i>pələṭ-iš</i>	> <i>pəltiš</i>	
8. Plpf. masc. sg.	<i>pəloṭ-elo</i>	<i>pələṭ-elo</i>	> <i>pəltelo</i>	
9. Pass. Fut.	<i>pəloṭ-a-še</i>	<i>pələṭ-a-še</i>	> <i>pəltaše</i>	
10. Caus. Fut.	<i>pəloṭ-av-še</i>	<i>pələṭ-av-še</i>	> <i>pəltavše</i>	

pəloṭ has a single form throughout the paradigm, whereas *pələṭ* shows variation between *pələṭ*- and *pəlṭ*-.

The ə-deletion does not operate on a monosyllabic base: *əḍ* ‘to touch’ has *əḍto* and *əḍiš* as present and future forms. The process takes place before a suffix that is -V(C) and is sensitive to phonetic as well as grammatical conditions. It does not take place when the suffix is a marker for case or number.³ For example, *əṭək* ‘to stop’ and *katər* ‘to cut’ have:

<i>əṭək-iš</i>	> <i>əṭkiš</i>	‘I will stop’
<i>əṭək-e</i>	> <i>əṭke</i>	‘he may stop’

3. A distinction between phonological and grammatical conditions is also maintained in the replacement of -o preceding a case marker except when the -o is a plural marker: *hatho-mā* > *hathamā* ‘in the handle’ but *hath-o-mā* > *hathomā* ‘in the hands’.

Pandit (1966: 241) infers that the innovation of plural -o in Guj. occurred historically after the -ə- deletion. He attributes this chronology for *rəməto* ‘games’ (no ə-deletion; -o pl.) and *rəmtō* ‘used to play’ (ə-deletion; -o masc. sg.). This chronology, however, would not explain *rəməte* ‘in the game’ (locative -e) and *rəmtē* ‘if he played’.

katər-iš > *katriš* 'I will cut'
katər-e > *katre* 'he may cut'

But *ətək* 'surname' and *katər* 'scissors' when combined with *-o* (pl.) or *-e* (instrumental/locative) leave the shwa of their second syllable intact.⁴

ətək-o > *ətəko* 'surnames'
ətək-e > *ətəke* 'by the surname'
katər-o > *katəro* 'scissors'
katər-e > *katəre* 'by the scissors'

This variation cannot be attributed to nominal or verbal bases since the *ə*-deletion operates in deriving an adjective from a noun: *aləs* 'laziness', *aləs-ū* > *alsū* 'lazy'; or one type of a noun from another type: *dərəd* 'disease': *dərəd-i* > *dərđi* 'patient'.

32.3.1.2. *a-Reduction and v-Insertion*

Guj. has a process of suffixation in which *-aš* or *-aṇ* combines with adjectives to create nouns.

(19)	Adjective	Noun
a. 'spicy'	<i>tik^h-</i>	<i>tik^haš</i>
b. 'sour'	<i>k^haṭ-</i>	<i>k^həṭaš</i>
c. 'unripe'	<i>kac-</i>	<i>kəcaš</i>
d. 'long'	<i>lamb-</i>	<i>ləmbaṇ</i>

This suffixation triggers another process: *-a-* of adjectival roots changes to *-ə-* in the corresponding nouns as exemplified in (19b–d). This process is also active in the derivation of pass. or caus. stems from verbs.

(20)	Fut.	Pass.Fut.	Caus.Fut.
a. 'play'	<i>rəm-še</i>	<i>rəm-a-še</i>	<i>rəm-aḍ-še</i>
b. 'walk'	<i>cal-še</i>	<i>cəl-a-še</i>	<i>cəl-aḍ-še</i>

4. Modi reports two plural forms for some words (*copḍio* and *copəḍio* from *copḍi* 'book') and considers them as representing two dialects differentiated by a presence vs. absence of a *ə*-deletion rule. She writes (Modi 1990: 206): "The uneducated speakers, or sometimes educated speakers in informal speech, insert vowel 'a' [i.e. *ə*]-thus adding a syllable with stress 'a'."

Such doublets, on the other hand, are an expected consequence of the analysis presented here. Given *copḍi-o* 'book+pl.', the application of the phonological conditioning to the base gives us *copḍio*. *copəḍio*, on the other hand, is the result of the grammatical condition (*-o* being a pl. marker) blocking the *ə*-deletion. If the data in fact represent two dialects, one can argue that when both conditions are simultaneously present, one dialect resolves the conflict by preferring phonological conditioning and the other dialect, grammatical conditioning.

c. 'run away'	<i>b^hag-še</i>	<i>b^həg-a-še</i>	<i>b^həg-aḍ-še</i>
d. 'drink'	<i>pi-še</i>	<i>pi-v-a-še</i>	<i>pi-v-aḍ-še</i>
e. 'sleep'	<i>su-še</i>	<i>su-v-a-še</i>	<i>su-v-aḍ-še</i>
f. 'eat'	<i>k^ha-še</i>	<i>k^hə-v-a-še</i>	<i>k^hə-v-aḍ-še</i>
g. 'spread'	<i>pat^hər-še</i>	<i>pət^hr-a-še</i>	<i>pət^hr-av-še</i>

As the forms in (20a) demonstrate, the pass. marker is *a* and the caus. marker is *-aḍ-* (with *-av-* as a variant). In (20b) and (20c) the *a*-reduction process has replaced *a* of the root with *ə* in pass. and caus. forms.

The forms in (20d) and (20e) show another process occurring under the same grammatical condition: *v*-insertion, with vowel-final verbs. The form in (20f) results from the application of both processes: *v*-insertion and *a*-reduction, whereas the form in (20g) is the consequence of *a*-reduction and *ə*-deletion.

32.3.2. Agreement

In addition to the aforementioned morphophonemic processes, Mistry (1992) notes the existence of a prosody-like phenomenon in Guj. that affects the phonetic composition of several constituents simultaneously. Compare (21) and (22).

- (21) *pelo ũco maṇəs g^həṇo moḍo ut^h-y-o*
 that tall man very late wake up
 'That tall man woke up very late.'

The *-o* in the demonstrative *pelo*, the adjective *ũco*, the intensifier *g^həṇo*, the adverb *moḍo*, and the verb *ut^hyo* in this sentence marks agreement with the masculine singular *maṇəs* 'man.'

- (22) *pela ũca maṇəst^hi g^həṇa moḍa ut^h-a-yū*
 that tall man-*t^hi* very late wake up-PASS.
 'That tall man could get up very late.'

Sentence (22) is a pass.-abilitative version of sentence (21) with the morphological marker *-a-* on the verb and *-t^hi* on the subject *maṇəs*. The word-final *-o* of the demonstrative, the adjective, the intensifier, and the adverb is replaced by *-a* since the agreement controller *maṇəs* has a marker *-t^hi*. This process, triggered by a marked agreement controller, instantiates the interfacing of syntax and phonology and is very pervasive in the language.

32.3.3. Intensification

Guj. contrasts single and geminated consonants: *cok^ha* ‘rice’ *cok^hk^ha* ‘clean’ (for additional examples, see § 32.5.3). Gemination additionally functions as a mechanism for intensification, as shown in (23).

- | | | | | |
|------|-------------------------|------------|--------------------------|-------------------|
| (23) | <i>k^haṭū</i> | ‘sour’ | <i>k^haṭṭū</i> | ‘very sour’ |
| | <i>sacū</i> | ‘true’ | <i>saccū</i> | ‘completely true’ |
| | <i>nəvū</i> | ‘new’ | <i>nəvvū</i> | ‘quite new’ |
| | <i>pakū</i> | ‘ripe’ | <i>pakkū</i> | ‘extremely ripe’ |
| | <i>same</i> | ‘in front’ | <i>samme</i> | ‘just in front’ |

Geminates occur mainly in adjectives and adverbs, and only these two classes allow intensification by gemination.

32.4. Internal differentiation

Ascertaining the sound system of a language poses a challenge since the phonetic repertoire of speakers varies. For some speakers, it could be restricted to their specific dialect, and for others it could encompass several dialects and languages at their command. Additionally, speakers show uneven fluency in languages and dialects other than their own. To sort out properties of the language, of the specific dialect(s), and of other languages becomes, therefore, a necessary but extremely difficult task.

Speakers show command over different varieties of the language, their preference being dependent on subtle sociolinguistic factors. A highly conspicuous division in Guj. is Sanskritic vs. non-Sanskritic styles, which are differentiated by various phonological and morphological features. A brief examination of compounding and clustering reveals some such features.

32.4.1. Compounding

Among the compounds with *mən* ‘mind’ as the first member, Guj. speakers use the following nouns (24) and adjectives (25).

- | | | |
|------|---|--------------------|
| (24) | a. <i>mən-o-mənt^hən</i> | ‘mental conflict’ |
| | b. <i>mən-meḷ</i> | ‘compatibility’ |
| | c. <i>mən-o-duk^h ~ mən-duk^h</i> | ‘mental agitation’ |
| | d. <i>mən-mənamṇa</i> | ‘appeasement’ |
| | e. <i>mən-murad</i> | ‘desire’ |
-
- | | | |
|------|----------------------|-------------|
| (25) | a. <i>mən-o-məy</i> | ‘mental’ |
| | b. <i>mən-pəsənd</i> | ‘agreeable’ |

- c. *mən-o-gəmyə* ~ *mən-gəmtū* 'pleasing'
 d. *mən-p^havtū* 'thoughtless'
 e. *mən-muk^htyar* 'uncontrolled, free'

Of these compounds, (24a) and (25a) have a stem-forming vowel *o* and are Sanskrit loans found in learned discourses; (24b) and (25b), without *-o-*, are compounds of everyday occurrence. (24c) and (25c) are instances of doublets with and without *-o-*. In (24d) and (25d) the second member is non-Sanskritic—structurally in (24) and historically in (25)—and the language has only the *-o*-less form. In (24e) and (25e) the second member is of Arabic origin, and again there is only a single form without the *-o-* in such instances. Thus, even in this restricted domain of compounds with *mən* as the first member, two styles show structural differences: the Sanskritic with a [[[*mən*]*o*]X] structure and the other with a [[*mən*]X] structure.

32.4.2. Consonant clusters

Clusters in Guj. are found primarily in loanwords. A Guj. speaker can thus be viewed as learning multiple systems through exposure to Skt., Perso-Arabic, and English loans, which affect and extend his native system. Words with initial clusters evidence these influences.

The only triconsonantal initial sequences are *str-*, *spr-*, and *smr-* found in the Skt. loans, such as *stri* 'woman', *spruha* 'desire', and *smruti* 'remembrance'. The gap *skr-* is filled by the English loans, e.g., *skru* 'screw'.

The biconsonantal initial sequences are of three types.

- i) with *s* as the first member and bilabial, dental, and velar voiceless stops as the second member: *skənd^h* 'shoulder', *sk^hələn* 'deviation', *stuti* 'prayer', *st^han* 'place', *spərd^ha* 'competition', and *sp^hurti* (~ *sfurti*) 'inspiration';
- ii) with *s* as the first member and *m*, *n*, *v*, *r* as a second member: *smərəṇ* 'memory', *sneh* 'affection', *svəjən* 'family member', *srot* 'stream';
- iii) with other consonants as the first member and *v*, *l*, *r*, *y* as a second member: *šyam* 'black', *šram* 'labor', *šlok* 'couplet', and *švas* 'breath'; or *pyalo* 'cup', *vrət* 'rite', *kleš* 'suffering', *tvəra* 'speed,' *jyoti* 'light'.

Clusters with *l* and *v* as the second member are quite constrained and have very low frequency. Only the Perso-Arabic loans (*myan* 'sheath' and *lyanət* 'disgrace') have clusters *my-* and *ly-*. Additionally, *gn* is the only word-initial

clustering of a stop and a nasal: *gnan* ‘knowledge’.⁵ The *ʃ* occurs in biconsonantal clusters either as a first member (discussed above) or as a second member in words used in learned discourse: *kṣəma* ‘forgiveness’, *kṣitiḥ* ‘horizon’. The first segment in such words is restricted to *k* and is articulated weakly by some and not at all by others. Among the retroflex sounds, only *ṭ* combines with word-initial *s* and occurs in English loans: *ṣṭeṣən* ‘station’, *ṣṭor* ‘store’.

The discussions of sibilants in § 32.2.2 and of compounding and clustering in this section show that there are occasional clues within a language to its internal differentiation.

32.5. Alternative analyses

In this section, murmur, aspiration, and gemination are reanalyzed in terms of their behavior in certain phonological processes.

32.5.1. *Murmur and aspiration*

In addition to the pan-Indian retroflex consonants that Guj. has, Pandit’s work shows the non-phonemic nature of vowel length, and the distinction of clear–murmur and open–close for vowels. Though murmur had been observed before, Pandit (1957) is an early treatise dealing with it. In this work, he attempted to show acoustic properties through spectrograms and kymograms, and proposed a specific analysis that considered consonantal aspiration and vocalic murmur allophonic. For him, (a) an aspirated consonant preceded by a clear vowel varying with an unaspirated consonant preceded by a murmured vowel (*lob^h* ~ *loh^h* ‘greed’) and (b) a murmured vowel varying with a clear one (*nəno* ~ *nano* ‘small’) are grounds for recognizing it as a separate entity. He considers *h* a source for murmur.

General works on Guj. (Cardona 1965, Vyas 1977), in their treatment of phonetics and phonemics, have incorporated Pandit’s observations and analysis, often even with the same examples. Though in-depth investigations of acoustic properties of murmur have appeared (Dave 1967, 1977, Fischer-Jørgensen 1967, Vyas 1979, Modi 1983, 1987), its source and analysis have not received the same scrutiny.

In Guj., the phonetic properties of words vary according to distinct, casual, or rapid speech. What is pronounced as three syllables *kəhe che* in formal speech is reduced to two syllables *kɛc^he* in casual speech and to a sin-

5. Some describe the first segment in this cluster as *j* rather than *g*. In my speech, the initial sound in *gnan* is similar to the initial sound in *gvalier* ‘a name of a city’, not *jvala* ‘blaze.’

gle syllable *kɛc^h* in rapid speech. Notice the deoralization of *h* and the emergence of a murmured vowel *ɛ* in casual and rapid speech. Additional instances of similar variations in formal vs. casual styles are listed in (26).

(26)	Formal	Casual	
a.	<i>hisab</i>	<i>ɪsab</i>	‘account’
b.	<i>humlo</i>	<i>ʊmlo</i>	‘attack’
c.	<i>hat^h</i>	<i>ət^h</i>	‘hand’
d.	<i>vəhelū</i>	<i>vɛlū</i>	‘early’
e.	<i>pəhoḷū</i>	<i>pɔḷū</i>	‘broad’
f.	<i>bəhar</i>	<i>bər⁶</i>	‘outside’
g.	<i>śahukar</i>	<i>śəvkar</i>	‘banker’
h.	<i>məhino</i>	<i>məyno</i>	‘month’

Three rules operate here.

1. A word-initial *hV* is realized as *V* (a murmured vowel),⁷ as in (26a–c).
2. A sequence of *əh* [–high vowel] is realized as [–high, more open, murmured vowel], as in (26d–f).
3. A sequence of *ə/ah* [+high vowel] is realized as murmured *ə/a* followed by a glide, as in (26g–h).

Conditions associated with these rules affect some *h*’s and keep other *h*’s intact, resulting in allomorphic variation for nouns (27) and verbs (28).

(27)	<i>əḱ</i>	‘privilege’	<i>na-həḱ</i>	> <i>nahəḱ</i>	‘wrongfully’
	<i>əl</i>	‘condition’	<i>be-hal</i>	> <i>behal</i>	‘bad condition’

(28)		<i>av-</i> ‘come’	<i>rəhe-</i> ‘stay’
	Fut. III	<i>av-ṣ-e</i>	<i>rɛ-ṣ-e</i>
	Fut. I sg.	<i>av-iṣ</i>	<i>rɛ-iṣ</i>
	Pres. II pl.	<i>av-o</i>	<i>rɔ</i>
	Pf.masc.sg.	<i>av-y-o</i>	<i>rəh-y-o</i>

(27) will have the following derivation.

(29)		<i>həḱ</i>	<i>nahəḱ</i>	<i>hal</i>	<i>behal</i>
	Rule 1	<i>əḱ</i>	—	<i>əl</i>	—
	Rule 2	—	—	—	—
	Rule 3	—	—	—	—
	Final forms	<i>əḱ</i>	<i>nahəḱ</i>	<i>əl</i>	<i>behal</i>

6. *a/ə* differ not only in the presence vs. absence of murmur, but the murmured *ə* is additionally more open than the clear *a*.

7. Some word-initial *h*’s remain intact in casual speech, indicating a conditioning factor for this rule. The constraint, however, is not clear at present.

The divergent allomorphy of *rəhe* ‘stay’ in (28) is the result of these three rules and an additional rule of vowel deletion that precedes them.

Rule A: A final root vowel is deleted before a suffix with a nonconsonantal initial segment.

The agglutinating morphology in the forms of *av* provides us the basis for setting up the underlying forms as in (30).

(30)	Fut.-III	Fut.-I-sg.	Pres.-II-pl.	Pf.-masc.-sg.
	<i>rəhe-še</i>	<i>rəhe-iš</i>	<i>rəhe-o</i>	<i>rəhe-yo</i>
Rule A	—	<i>rəh-iš</i>	<i>rəh-o</i>	<i>rəh-yo</i>
Rule 2	<i>rɛ-še</i>	—	<i>rɔ</i>	—
Rule 3	—	<i>rɔiš</i>	—	—
Final	<i>rɛše</i>	<i>rɔiš</i>	<i>rɔ</i>	<i>rəhyo</i>

Rule 1 accounts for murmured vowels in word-initial position and rules 2 and 3 for *ɛ*, *ɔ*, *a*, *əi*, *ai*, *əu*, *au* in other positions. The existence of allomorphs with *h* shows that not all *h*’s become murmur.

In addition to *h* in certain allomorphs, *h* remains invariable in certain words. The loans from Skt. *hastā* ‘hand’ and *harsā* ‘joy’ have only one phonetic realization but their derivatives, *hat^h* and *hərək^h* respectively, are also pronounced as *at^h* and *ərək^h*. The following are some of the words with an invariable *h*.

(31)	<i>hani</i>	‘damage’	<i>rəhəsyā</i>	‘secret’
	<i>hīnd</i>	‘India’	<i>jaherat</i>	‘advertisement’
	<i>mohən</i>	‘personal name’	<i>nihaḷvū</i>	‘to watch’

Additionally, Guj. has a set of contrasting words *kəro* ‘do! (imperative)’ and *kəro* ‘wall’, where the word-medial murmured *ə* cannot be derived by Rule 2 or Rule 3, nor can one consider *h* to be its source. Our investigation thus shows that not all *h*’s become murmured and not all instances of murmur derive from *h*.

Turning to the analysis, the grouping of aspiration and murmur into a single phoneme (Pandit 1957: 169–71, Cardona 1965: 29–30) implies that aspirated stops and murmured vowels are clusters and complex nuclei, respectively. As illustrated in (32)–(34), these sounds behave as single segments in three rules: *ə*-deletion, *a*-reduction, and *v*-insertion, discussed in § 32.3.1.1 and § 32.3.1.2.

(32)	Verb+Fut.	Verb+Pres.
a. 'harass'	<i>pəjəv-še</i>	<i>pəjv-e</i>
b. 'describe'	<i>vəṁṇəv-še</i>	<i>vəṁṇəv-e</i>
c. 'arrange'	<i>goʈʰəv-še</i>	<i>goʈʰv-e</i>
(33)	Verb+Pf.	Verb+Pass.+Pf.
a. 'use'	<i>vapər-yo</i>	<i>vəpr-a-yo</i>
b. 'attract'	<i>akərš-yo</i>	<i>akərš-a-yo</i>
c. 'recognize'	<i>parəkʰ-yo</i>	<i>pərəkʰ-a-yo</i>
(34)	Verb+Impf.	Verb+Pass.+Impf.
a. 'see'	<i>jo-to</i>	<i>jo-v-a-tū</i>
b. 'sit'	<i>bes-to</i>	<i>bes-a-tū</i>
c. 'say'	<i>keh-to</i>	<i>keh-v-a-tū</i>

For the ə-deletion, the *tʰ* in (32c) patterns with the *j* (a single segment) in (32a) and not with the *ṁ* (a cluster) in 32b. The *kʰ* in (33c) shows the same characteristic in the *a*-reduction process by being similar to the *r* in (33a). Similarly, in the *v*-insertion process, the *-eh* (murmur vowel) in (34c) and the *-o* in (34a) behave alike. There is thus a more general account of phonology possible if the aspirated stops and the murmured vowels are analyzed as single segments.

The treatment of the dual aspects of murmur—its phonetic characteristics and distribution—is particularly notable in Modi's work (1983, 1987). For her, "murmur, aspiration, nasalization and open mid-vowels are the main trouble spots of Gujarati phonology" (Modi 1989: 17). She reinterprets Fischer-Jørgensen's phonetic study of murmur, and supplements that with her own studies based on tomograms and X-ray pictures. She argues for the presence of two separate phenomena distributed dialectally: tight phonation in peninsular Guj. and murmur in the mainland dialect. Murmur, for her, is due to a lowered and slightly spread status of the glottis and it occurs in defined linguistic contexts. Tight phonation, on the other hand, is attributed to a raised glottis with tension in the surrounding muscles and is present throughout an utterance. She moreover suggests that murmur and aspiration be treated as a prosody, spread over several segments. A fuller account of Guj. phonology along these lines is a desideratum.

32.5.2. Gemination

Guj. has geminates—stops, fricatives, nasals, voiced, voiceless, aspirated, unaspirated—as shown in (35).

(35)	<i>sikko</i>	‘coin’	<i>d^humməs</i>	‘mist’
	<i>dəbbo</i>	‘box’	<i>unnəti</i>	‘prosperity’
	<i>pəɽ^ht^hər</i>	‘stone’	<i>c^həṇṇu</i> ⁸	‘ninety six’
	<i>b^həɽ^ht^hi</i>	‘furnace’	<i>təllin</i>	‘absorbed’
	<i>lucco</i>	‘cunning’	<i>kisso</i>	‘episode’

Geminates contrast with simple consonants: *muko* ‘put!’ vs. *mukko* ‘fist’, *aḍo* ‘touch!’ vs. *aḍḍo* ‘meeting place’, and *bəne* ‘may happen’ vs. *bəṇṇe* ‘both’.

The previous descriptions (Pandit 1961: 51, Cardona 1965: 31) treat geminates as long consonants. There are reasons for considering the possibility that geminates should be analyzed as clusters of two identical segments. For one thing, contrasting *ucaɽ* ‘anxiety’ and *uccar* ‘pronunciation’, the *u* in *ucaɽ* is a lax sound prevalent in an open syllable, whereas the one followed by a geminate in *uccar* is similar to the *u* in *udgar* ‘utterance’ in showing tenseness typical in a closed syllable.

Moreover, the *ə*-deletion process (see § 32.3.1.1) is restricted to CəCV# sequences and does not apply to CəCCV#. Guj. thus has *avərda* ‘life’, *gəṇṇətri* ‘calculation’, and *kotəṇi* ‘carving’, but no words with a -CəCV# sequence. If geminates were to be considered single segments, such words as *apətti* ‘calamity’ and *təməṇṇa* ‘desire’ would be the only exceptions to the above generalization. Alternately, geminates analyzed as a cluster of two identical consonants make the description of the syllable structure of words and the *ə*-deletion process more systematic and general.

32.6. Southern Gujarati

The available studies on Guj. are based primarily on speakers from either the peninsular or the northern part of Gujarat. Consequently, some of the reported characteristics do not hold for the southern Guj. that I speak. For example, the deaspiration in cases such as *lək^h-to* > *ləkto* ‘writing’ that Cardona (1965: 49) mentions is not true for my speech. I have a contrast between *pəḍ-to* > *pəḍto* ‘falling’ and *pəḍ^h-to* > *pəḍ^hto* ‘studying’; as well as *lad-še* > *ladše*

8. Concerning geminate *ṇ*, Pandit (1957: 165) states: “Save for one loan-word from Marathi, *əṇṇa* ‘elder brother’ current mainly among the Gujarati speaking Marathis, *ṇ* does not occur as a long-geminated consonant.” This shows that the native word *c^həṇṇu* ‘ninety-six’ and -*ṇṇ*- as a variant of -*m*- in words such as *baṇṇu* (~ *baṇu*) ‘door’ or *gəṇṇi* (~ *gəṇi*) ‘strainer’ have escaped his attention. Modi (1989: 18) also draws attention to Pandit’s omission of data that led to the unnecessarily restricted statement about the distribution of palatal and velar nasals. This is surprising since Pandit’s description in other respects is an account of an optimal system which some Guj. speakers command.

‘will load’ and *lad^h-še* > *lad^hše* ‘will gain’. The region even has a word *b^hṛavū* ‘to die’ in which a murmured vowel is preceded by a voiced aspirate, countering the claim that only a clear vowel occurs after aspirated consonants (Pandit 1957: 169).

Southern Guj. also shows some differences in nasalization. In standard Guj. nasalized and clear vowels contrast in word-final position: *vəṣmā* ‘under control’ and *cəṣma* ‘glasses’. Even a grammatical difference is conveyed through nasalization. For example, an agreeing adjective modifying a plural noun ends in *-a* with masculine nouns and *-ā* with neuter nouns: *vəsma divəso* (masc. pl.) ‘difficult days’ and *vəsmā vərəso* (neu. pl.) ‘difficult years’. In Southern Guj., however, there is no audible difference between the last vowels in these examples.⁹ In this dialect even monosyllabic words such as *ṣū* ‘what’ and *hū* ‘I’ are pronounced as *ṣu* and *hu* in one register and *ḥu* and *ṽ* in another. Moreover, unlike standard Guj., southern Guj. lacks nasalized *ə*. *səṣsar* ‘world’ and *səvad* ‘dialog’ are *sənsar* and *səmvad* in my speech.

We thus find that southern Guj., while it exhibits murmur, aspiration, and nasalization, does vary in the prominence of these properties. A cursory comparison indicates maximum murmuring and minimum nasality in the south.

32.7. Conclusion

This chapter has presented a brief outline of the sound structure of Guj. and its connection with other levels of the language, and has suggested an alternative analysis for murmur, aspiration, and gemination. The outline is restricted to segmental properties and reports observations and analyses largely from the existing literature. The narration of instances of interaction of phonology with morphology, syntax, and semantics and characteristics that differentiate two styles of the language, though new, is merely suggestive. Moreover, a difference in perspective—the concern with phonological processes—is at the root of the suggested alternative analysis. The data and characteristics of southern Guj. are an added facet of this study.

9. In my speech, nasalization in a final position is present in a couple of cases: *tyā* ‘there’ and *kyā* ‘where’ end with nasalized vowels contrasting with *-tya* in *pətya* ‘completed’ and *-kya* in *ṭəkyā* ‘lasted’. For an informative account of variation in nasalization see Pandit (1972: 57–80), where the phenomenon is examined in terms of three factors: education, locality, and ethnicity.

References

- Bloch, Jules. 1965. *Indo-Aryan from the Vedas to Modern Times*, translated by A. Master. Paris: Adrien-Maisonneuve.
- Cardona, George. 1965. *A Gujarati Reference Grammar*. Philadelphia: University of Pennsylvania Press.
- Chatterji, S. K. 1926. *The Origin and Development of the Bengali Language*. Calcutta: Calcutta University Press.
- Dave, Radhekanth. 1967. "A Formant Analysis of the Clear, Nasalized, and Murmured Vowels in Gujarati." *Indian Linguistics* 28:1–30.
- . 1977. *Studies in Gujarati Phonology and Phonetics*. Ph.D. dissertation, Cornell University.
- Dave, T. N. 1932. "Notes on Gujarati Phonology." *Bulletin of the School of Oriental Studies* 6: 673–78.
- . 1948. *The Language of Maha-Gujarat*. Bombay: Gujarat Research Society.
- Firth, J. R. 1957. "Phonetic Observations on Gujarati." *Bulletin of the School of Oriental and African Studies* 20:231–41.
- Fischer-Jørgensen, Eli. 1967. "Phonetic Analysis of Breathy (Murmured) Vowels." *Indian Linguistics* 28: 71–139.
- Grierson, George A. 1928. "Gujarati." In *Linguistic Survey of India*, vol. 9, pt. 2, pp. 323–477. Calcutta: Government of India.
- Master, Alfred. 1925. "Stress Accent in Modern Gujarati." *Journal of the Bombay Branch of the Royal Asiatic Society*, new series 1: 76–94.
- Mistry, P. J. 1992. "Objecthood and Specificity in Gujarati." In *Agreement in South Asian Languages*, ed. Manindra Verma and Alice Davison, to appear.
- Modi, Bharati. 1983. *Some Issues in the Phonology of Gujarati*. Ph.D. dissertation, Maharaja Sayajirao University, Baroda.
- . 1987. "Rethinking of 'Murmur in Gujarati'." *Indian Linguistics* 47: 39–55.
- . 1989. "A Critical Survey of Phonological Studies in Gujarati." In *National Seminar on Indian Linguistics: An Overview*. Baroda: Maharaja Sayajirao University.
- . 1990. "The Relationship between Phonology and Morphology: Morphology Phonologized." *Journal of the Oriental Institute*, Baroda. 38: 291–303.
- Pandit, P. B. 1955. 'E and O in Gujarati.' *Indian Linguistics* 14: 36–44.

- . 1957. "Nasalization, Aspiration and Murmur in Gujarati." *Indian Linguistics* 17: 165–222.
- . 1958. "Duration, Syllable and Juncture in Gujarati." *Indian Linguistics. Turner Jubilee Volume* 1: 212–19.
- . 1961. "Historical Phonology of Gujarati Vowels." *Language* 37: 54–66.
- . 1966. *Gujarāti bhāshānū dhvani-svarup ane dhvani-parivartan* (Sound system and sound changes in Gujarati). Ahmedabad: Gujarat University.
- . 1972. *India as a Sociolinguistic Area*. Poona: University of Poona.
- . 1973. "Gujarāti bhāshānā vidhāyak bālo" (Historical evolution of Gujarati). In *Gujarāti sāhityano itihās* (A history of Gujarati literature), ed. U. Joshi, A. Rawal, and Y. Shukla. Ahmedabad: Gujarati Sahitya Parishad.
- Turner, R. L. 1915. "The Indo-Aryan Nasals in Gujarati." *Journal of the Royal Asiatic Society* 1033–38.
- . 1921. "Gujarati Phonology." *Journal of the Royal Asiatic Society* 329–65, 505–44.
- . 1925. "The e and o Vowels in Gujarati." *Sir Asutosh Mookerjee Silver Jubilee Volumes*, vol. 3, *Orientalia*, pt. 2, pp. 337–47. Calcutta.
- . 1966. *A Comparative Dictionary of the Indo-Aryan Languages*. London: Oxford University Press.
- Vyas, Manorama. 1979. *Phonation Types in Gujarati*. M.Phil. dissertation, University of London.
- Vyas, Yogendra. 1977. *Gujarāti bhāshānū vyākaraṇ* (A grammar of the Gujarati language). Ahmedabad: Sahitya Mudranalaya.

This page intentionally left blank.

Persian Phonology

Gernot L. Windfuhr

University of Michigan

33.1. Introduction

Persian is the official language of Iran and Tajikistan, and one of the official languages of Afghanistan, beside Pashto. The total number of speakers may be estimated between 70 and 80 millions.

From the 10th until the 17th century, Persian was the language of the dominant political and cultural power in the Near and Middle East outside Arab countries, and as such was the dominant lingua franca and language of literature in an area stretching from Iran into Southern Russia, Central Asia, and India. The Moguls of Dehli elected Persian as their administrative and cultural language for their multilingual and multi-ethnic domains, until English gradually replaced it. Sultans of the Ottoman Empire are known to have composed in Persian. As a result, both Ottoman Turkish and Hindi/Urdu are heavily Persianized.

Persian emerged by the 8th century A.D. out of Middle Persian after the coming of Islam, first in the East, after which the centers moved to Iran proper. European merchants and missionaries in Iran and employees of the British East India Company in India had to learn Persian, and produced the first European grammars of Persian beginning in the 17th century.

Persian is an Indo-European language. The name Persian derives from the Greek *Persis*. It refers to the SW province of the Iranian plateau, where the Achaemenids established their empire by the middle of the first millennium B.C. Its Old Iranian form was *Parsa*, its Arabicized form is *Fars*, thus *Farsi* as the name of the language. The name Iran derives from Old Iranian **aryānām khshathra* ‘the realm of the Aryans’, later *Iranshahr*, then *Iran*.

33.2. Classification

Diachronically, Iranian languages are divided into Old (up to the 3rd century B.C.), including Old Persian and Avestan, the language of Zarathushtra; Middle (up to the 7th century. A.D., some lasting into the 13th and 14th

centuries), including Middle Persian, Parthian, Sogdian, Saka, Khwarezmian, and Bactrian; and New Iranian. Geographically, they are grouped into Eastern and Western, with a northern and southern group. That grouping should be revised to Northern and Southern:

1) A northern (not eastern) group, including Ossetic in the central Caucasus; Pashto, the main language of Afghanistan; other languages in and about Afghanistan, including the Pamir languages; and Yaghnobi in southern Central Asia, the successor of Middle Iranian Sogdian, which once stretched to the borders of China.

2) A southern (not western) group found in Iran, Iraq, Syria, Turkey, and the southern Caucasus. It includes Baluchi found from eastern Iran and western Pakistan to Central Asia; Kurdish, found in western Iran, northern Iraq, Syria, eastern Turkey, Armenia; Gurani dialects found in the middle Zagros near the Iraqi border and near Mosul in Iraq; Dimili (Zaza) in eastern Turkey; Taleshi and Tati dialects stretching from Azerbaijan to the Central dialects; Alborz dialects, and the Caspian dialects Gilaki and Mazanderani; the Fars dialects, the Lar dialects, and Bashkardi along the Gulf; and Kumzari in Masandam in Oman.

33.3. Phonology

A major factor in the development of Persian phonology is the impact of the Arabic loan component, which constitutes approximately 50% of the lexicon, with approximately 25% frequency of usage. Also, the effects of social dynamics are difficult to determine. A particularly challenging aspect is the rather unique cultural setting in Iran, where schooling involves intensive study of classical literature and prosody through all levels of education, and where the ideal of classical Persian continues to be an integral part of the culture to a degree unknown elsewhere.

The phonological system posited for Persian is the result of many approaches, each with preconceived assumptions about such systems, based on a combination of impressionistic and experimental phonetic analyses, acoustic or auditory, refracted by fuzzy notions and definitions about social and geographic dialects, and mostly incomplete. As such, the study of Persian phonology is not different from the study of other phonologies.

Overall, there has been relatively little experimental phonetic research on Persian, most of which happens to have been conducted by Soviet scholars. In particular, the hierarchies of immediate adjacency, syllable structure, boundaries, and stress need more thorough investigation and definition.

It should be noted that the classification by binary distinctive opposition is a descriptive device for what in actuality are hierarchies of unequal pairs, which determine their dynamics and dialectic.

33.3.1. Segmental phonology

Contemporary Iranian Persian has 29 phonemes. By point and manner of articulation the consonants are identified as in Table 33-1.

Table 33-1. Persian Consonants

	Bilabial	Labio-Dental	Dental-Alveolar	Alveo-Palatal	Velar	Post-Velar	Glottal
Stops							
Voiceless	p		t		k		ʔ
Voiced	b		d		g		
Affricates							
Voiceless				č			
Voiced				ǰ			
Fricatives							
Voiceless		f	s	š		x	h
Voiced		v	z	ž		q	
Liquids			r, l				
Nasals	m		n				
Semi-Vowels				y			

33.3.1.1. Vowels

Prior to experimental studies, there were problems with the pair *a* vs. *ā*. Observations focused on the closeness of *a* to *e*, and the various degrees of the rounding of *ā*, which tends to be quite noticeable in two seemingly unrelated social contexts: poetic diction and recitation, and to a lesser degree, in female speech.

The articulatory features of the six vowels were established through phonetic experiment by Gaprindashvili and Giunashvili (1964). They demonstrated that *a* and *ā*, as well as *o*, are mid-level, whereas *e* is located between *i* and *a*; and that *a* is central, whereas *ā* tends to oscillate between central and back position (Table 33-2).

Table 33-2. Persian Vowels

	front	central	back
high	i		u
mid-high		e	
mid		a ā	o

The classificatory relationships may be shown as in Table 33-3.

Table 33-3. Persian Consonant Classification

		labial	dental	palatal	velar
Stops	voiceless	p	t	č	k
	voiced	b	d	ǰ	g
Fricatives	voiceless	f	s	š	x
	voiced	v	z	ž	q
Nasals		m	n		
Liquids			l, r		

h, *y*, and *ʔ* as glides are aligned with the vowels as in Table 33-4.

Table 33-4. Persian Glides and Vowels

		low	front	back
Glides		h	y	ʔ
Vowels	long	ā	i	u
	short	a	e	o

33.3.2. Vowels

33.3.2.1. Minimal pairs

o vs. *u*

<i>pol</i>	‘bridge’	<i>xod</i>	‘self, own’	<i>to</i>	‘thou, you’
<i>pul</i>	‘money’	<i>xud</i>	‘helmet’	<i>tu</i>	‘inside’

o vs. *ā*

<i>dom</i>	‘tail’	<i>šod</i>	‘happened’	<i>to</i>	‘thou, you’
<i>dām</i>	‘net’	<i>šād</i>	‘happy’	<i>tā</i>	‘till’

a vs. *a*

<i>kam</i>	'little'	<i>nam</i>	'wet'	<i>kar</i>	'deaf'
<i>kām</i>	'desire'	<i>nām</i>	'name'	<i>kār</i>	'work'

a vs. *e*

<i>dah</i>	'ten'	<i>sar</i>	'head'	<i>motanavvār</i>	'enlightened'
<i>deh</i>	'village'	<i>serr</i>	'secret'	<i>motanavvér</i>	'enlightening'

e vs. *i*

<i>serr</i>	'secret'	<i>se</i>	'3'	<i>ke</i>	'that'
<i>sir</i>	'garlic'	<i>si</i>	'30'	<i>ki</i>	'who'

33.3.2.2. *Allophonic variation*

Vocalic length is affected by stress. Length distinction in contemporary Iranian Persian is neutralized in most environments. It is retained in two environments. One is open, non-final pretonic syllables, established by the phonetic experiments of Sokolova et al. (1952: 178–92); e.g., in *se-dā* 'voice' vs. *se:f-tār* 'harder'. Sokolova therefore suggested that the basic distinction is that between relative stability, with the opposition of stable vs. non-stable vowels, instead of long vs. short. The other is initial stressed syllables, as shown by Windfuhr (1979: 135–37), where the distinction is long vs. short:

<i>ā/a</i>	<i>nā́-kas</i>	<i>ná-gu</i>	'nobody!'	'don't say!'
<i>i/e</i>	<i>bí-adab</i>	<i>bé-gu</i>	'impolite!'	'say!'
<i>u/o</i>	<i>húšang</i>	<i>hóseyñ</i>	'Hushang!'	'Hoseyn!'

Given that basic features tend to be posited on the basis of opposition in stressed position, length appears to be the primary distinctive feature.

The unstable short vowels *a*, *e*, *o* show a certain degree of assimilation in height, rounding, and length. There is certainly a continuum between social register and dialect register.

1) *e* and *o* tend to be raised before *i* as in *kelíd* > *kilíd* 'key', *qeirát* 'zeal', and/or rounded, as in *békon* > *békon/bókon* 'do!', note the same in the loan *belúz* > *bolúz/bulúz* 'blouse'; *bolúr* > *bulúr* 'crystal'. 2) *a* tends to be raised to *e* in the context of sibilants; e.g., *madrasé* > *madresé* 'school', note the frequent assimilation in the personal suffix *-aš*, as in *ketāb-aš* > *ketāb-eš* 'his book', also found, by analogy, in *-at*, as in *ketāb-at* > *ketāb-et* 'your book' (but never 1sg. suffix *-am* > **-em*). An instructive case is the assimilation of *čašm* > *češm* 'eye'. Such did not occur in *čašm* 'all right, will be done!' (from *be nūr-e čašm-am* 'by the light of my eyes'), which is no longer recognized as 'eye'. There is also length assimilation, most noticeable in *a*; e.g., *nahār* > *nāhār* 'lunch', some dictionaries listing both forms. 3) *i* tends to be shortened before *y*; e.g., *xiyābān* 'street', *piyālē* 'pitcher', *siyāsāt* 'politics'.

Note: The oldest suggestion for the basic difference between the two series of vowels is length distinction. The basic criterion cited is diachronic phonology. Vocalic length distinction is continued from Old, Middle, and Early New Persian. It is also argued that the same distinction is reflected in the representation of the vowels in Arabic script, where the inherited vocalic length of Persian words is represented by *alif*, *yā*, and *wāw*, respectively, whereas short vowels are represented by the diacritics *fathe*, *kasre*, and *damme*. The representation of the vocalic system of Persian thus follows that of the Arabic loan component. A further argument comes from prosody, where vowels identified as long count as long, and short as short.

The instability of *a*, *e*, and *o* is reflected in the earliest European grammars of Persian of the 17th century.

33.3.3. Consonants

33.3.3.1. Stops

The two series of stops *p t k* and *b d g*, together with the alveopalatals *č ĵ*, are generally classified into voiceless vs. voiced stops.

p vs. *b*

<i>pā</i>	‘foot’	<i>pand</i>	‘advice’	<i>pud</i>	‘woof [in weaving]’
<i>bā</i>	‘with’	<i>band</i>	‘band’	<i>bud</i>	‘was’

t vs. *d*

<i>tār</i>	‘string’	<i>tur</i>	‘net’	<i>tir</i>	‘arrow’
<i>dār</i>	‘gallows’	<i>dur</i>	‘far’	<i>dir</i>	‘late’

č vs. *ĵ*

<i>čub</i>	‘wood’	<i>čang</i>	‘claw’	<i>čost</i>	‘quick’
<i>ĵub</i>	‘stream’	<i>ĵang</i>	‘war’	<i>ĵost</i>	‘searched’

k vs. *g*

<i>kām</i>	‘palate’	<i>kur</i>	‘blind’	<i>kir</i>	‘penis’
<i>gām</i>	‘step’	<i>gur</i>	‘grave’	<i>gir</i>	‘catch, hold’

33.3.3.1.1. ALLOPHONIC VARIATION

A major question is the conditioned variation of consonants. The three general environments to be considered are 1) initial and prevocalic position, 2) postvocalic and final position, and 3) clusters.

Environments 1 and 2: *p t č k* are always voiceless, and aspirated; *b d ĵ g* are never aspirated; e.g., *t^har* ‘wet’ but *dar* ‘in’. Environment 3: *p t č k* retain aspirate release; *b d ĵ g* are generally devoiced; e.g., *māt^h* ‘perplexed’ and *māč*

'Media'. Environment 3: The distinction between the two series is neutralized; e.g., *kaftár* > *kafḏár* 'pigeon'. (This effect is occasionally reflected in writing; e.g., *moškel* and *mošgel* 'difficult'.)

While devoicing and neutralization suggest the distinction of fortis vs. lenis as a factor in distinguishing the two stop series, they are conditioned, and thus secondary. Accordingly, voice distinction is the primary feature.

Velars. As for vocalic environment, velars tend to be fronted before front vowels, including *a*; e.g., *k'eyf* 'enjoyment', *k'ar* 'deaf', *g'ij* 'dizzy'. They tend to be backed before back vowels; e.g., *kur* 'blind', *gur* 'grave'.

Note: Fronting is part of a major cross-linguistic isogloss that begins in central Iran and stretches across Azerbaijan and the Caucasus.

33.3.3.1.2. GEMINATION

Gemination is phonemic; e.g., *korré* 'foal' vs. *koré* 'globe', *darré* 'valley' vs. *daré* 'the door', *zann* 'opinion' vs. *zan* 'woman, wife'.

33.3.3.1.3. ALLOPHONIC VARIATION

The phonetic realization of gemination appears to be confined to formal registers, but even there appears to be rare; e.g., both *korré* 'foal' and *koré* 'globe' are mostly pronounced non-geminate. Distinct gemination is occasionally provided by native speakers in cases of potential ambiguity, which are few; e.g., *bannā* 'architect' vs. *banā* 'building'. (In writing, the diacritic *tašdid* is added in cases of possible ambiguity.)

Because of the speaker's ability to foreground gemination, and because gemination also surfaces in poetic diction, it could be argued that gemination is an underlying feature of Iranian Persian.

Note: Gemination is also posited on diachronic grounds, particularly with respect to the Arabic loan component.

33.3.3.2. Fricatives

The fricative series *f s š x* and *v z ž q* are distinguished by voice.

f vs. *v*

<i>fām</i>	'tint'	<i>nāf</i>	'navel'	<i>kāfé</i>	'cafe'
<i>vām</i>	'loan'	<i>nāv</i>	'boat, pipe'	<i>kāvé</i>	'Kave'

s vs. *z*

<i>sir</i>	'satisfied'	<i>sard</i>	'cold'	<i>māsé</i>	'fine sand'
<i>zir</i>	'beneath'	<i>zard</i>	'yellow'	<i>māzé</i>	'spine'

š vs. *ž*

<i>šāš</i>	'piss'	<i>ašur</i>	'Assyria'	<i>guš</i>	'ear'
<i>žāž</i>	'idle talk'	<i>a-žur</i>	'up to date'	<i>guž</i>	'haunch'

x vs. *q*

<i>xār</i>	‘thorn’	<i>mox</i>	‘brain’	<i>faxr</i>	‘pride’
<i>qār</i>	‘cave’	<i>moq</i>	‘magus’	<i>faqr</i>	‘poverty’

The status of *q*. The characteristics of *q* remain the most problematic issue of the phonemic system of Iranian Persian. Experimental phonetic studies by Sokolova and Gaprindashvili and Guinashvili affirmed the impression of considerable conditioned variation, varying itself with the speaker.

The first point to note is that the critical conditioning environments for *q* are the same as those for the stops, but with divergent effects. 1) In prevocalic initial position, *q* tends to be a voiced stop, often followed by a fricative release. 2) In final position, it is a devoiced stop, or fricative, both with equal frequency. 3) In clusters, voicing is neutralized. Most prominently, it tends to be devoiced before *t*; e.g., *eqtesād* ‘economics’. The shift to *xt* is complete in the word *vaqt* > *vaxt* ‘time’, probably due to the functional development of *vāqt-i* ‘at a time’ to the conjunction *vāxti* ‘when’. In addition, in intervocalic position *q* is a voiced fricative, which supports its systemic pairing with *x*.

Note: Diachronically, *q* is the result of the confluence of 1) the Middle Persian voiced velar fricative, 2) Arabic voiced fricative *gheyn* as well as unvoiced *qāf*, 3) Turkish *g* and *k* in the position before low vowels. The presently generally used symbol for this phoneme, *q*, is a compromise.

33.3.4. Liquids and nasals

The liquids *l* and *r* are strictly dental, and unrounded.

r vs. *l*

<i>xor</i>	‘sun’	<i>mār</i>	‘snake’	<i>lām</i>	‘L’
<i>xol</i>	‘stupid’	<i>māl</i>	‘property’	<i>rām</i>	‘quiet’

n vs. *m*

<i>nār</i>	‘fire’	<i>xāné</i>	‘house’	<i>ney</i>	‘flute’
<i>mār</i>	‘snake’	<i>xāmé</i>	‘cream’	<i>mey</i>	‘wine’

Of the two nasals, *n* assimilates to velar; e.g., backed velar after long *ā*; e.g., *bāng* ‘shout’; fronted velar after short *a*; e.g., *rang* ‘color’.

Allophonic devoicing occurs in the environment C-#; e.g., *saṭtʃ*# ‘pail’, *faxtʃ*# ‘pride’; *matʃ*# ‘text’, *hokʃ*# ‘order’.

Note: There is occasional rounding of *l* in the name of *allāh*, and the adhortative call *yāllā* ‘let’s go’.

33.3.5. Glides

The glides *h*, *y*, and *ʔ* are phonemic, and show no distributional restrictions; e.g., *har* ‘each’, *nahā́r* ‘lunch’, *dah* ‘10’; *yād* ‘memory’, *ā́yā* ‘whether’, *čā́y* ‘tea’; *ʔā́ftāb* ‘sunshine’, *sāʔāt* ‘hour, watch’, *nowʔ* ‘kind (of)’.

Note: Diachronically, the glottal stop is the result of the confluence of the non-phonemic vocalic onset in the Persian component, and of both *ʔyn* and *hamze* in the Arabic loan component.

33.3.5.1. Allophonic variation

ʔ and *h* show considerable conditioned variation. They remain unchanged in initial position, but have the following realizations: 1) Intervocally, they tend to be articulated fleetingly. 2) In postvocalic preconsonantal position they result in compensatory lengthening. 3) In final, prepausal position, they are not articulated in the absence of a following vowel; e.g., *šeʔr* > *še:r* ‘poem, poetry’, *šahr* > *ša:r* ‘city’, and final *nowʔ* > *now* ‘kind, sort’, *šarh* > *šar* ‘description’.

33.4. Syllabic structure

The syllabic structure of Iranian Persian is [C1]-[V]-[C2]. [C1] is always a single consonant, while [C2] may be one or two consonants, or pause. [V] may be long or short.

As to sequential syllables, it follows that any consonant preceding a vowel must be the initial consonant of the next syllable; e.g., morphemic *dāst-am*, but syllabic *dās-tam* ‘my hand’; morphemic *jamʔ-aš*, but syllabic *jam-ʔaš* ‘its plural form’. Note that *h* and *ʔ* are retained in this position.

33.4.1. Diphthongs

It has been argued that *ey* and *ow* represent phonemic units; e.g., *keyk* ‘flea’, *dowr* ‘(a)round’, and with pause, *key#* ‘when’, *dow#* ‘run’. Syllabic structure rules clarify the issue: *ey* and *ow* are never followed by two consonants, and must therefore be analyzed as vowel + consonant, the latter being part of [C2], so that *keyk* and *dowr* have the structure CVCC.

33.4.1.1. Semivowel *y*, and *v/w*

Neither *y* nor *v* has distributional restrictions; e.g., *yād* ‘memory’, *ā́yā* ‘whether’, *čā́y* ‘tea’; *vām* ‘loan’, *āvārd* ‘brought’, *dāv-talāb* ‘applicant’. However, *a* + *v* changes to allophonic *ow*.

The sequence *ow* contrasts with the sequence *ov*; e.g., *gerów* ‘(financial) deposit’ vs. *nomóvv* ‘growth’. The most plausible solution to this problem is to posit the underlying sequence *av* > *ow* vs. *ov* > *ov*. (In writing, *ey* and *ow* sequences used to be marked by the superscript *fathe*, but today are marked by *kasre* for *ey* and by *damme* for *ow*.)

Note: The diachronic argument for the analysis *ey* and *ow* as vowel + consonant is based on the fact that the two originate in earlier *ay* and *aw*; e.g., Iranian *kay* > *key* ‘when’ and *dav* > *dow* ‘(a) run’; Arabic *dawr* > *dowr* ‘around, turn’ and *dayr* > *deyr* ‘monastery’.

33.4.2. Hiatus

No two adjacent vowels are allowed. The hiatus is marked by a glide which is realized as a fleeting glottal stop or palatal glide, depending on the stress pattern; e.g., *mí-gu-ʔi* or *mí-gu-yi* ‘you are saying’. Phonemically it may be correlated with either *y* or *ʔ*.

Note: The absence of hiatus is the single suprasegmental feature, besides stress shifts by contraction, that distinguishes less formal registers; e.g., *ket-āb-hā-y-am* vs. *ketāb-hā-m* ‘book-pl.-mine’.

33.5. Stress

Stress in Persian is dynamic, predictable, and non-phonemic. Word stress is rightmost, either the base word or the derivational suffix; e.g., *dast* ‘hand’, *dast-é* ‘handle, bunch’, *mard-āné* ‘man-ly, men’s’, *sabz-í* ‘greenness, vegetable’. The same applies to verbal participles; e.g., *oftād* ‘fell’, *oftād-é* ‘fallen’. Stress is likewise on the plural marker; e.g., *ketāb-hā* ‘books’, *digar-ān* ‘the other persons’. In extended noun phrases primary stress is on the final stressable syllable; e.g., *javāb-e xūb* ‘a good answer’, *kešvār-e irān* ‘the country of Iran’.

Phrase stress is leftmost and phrase-initial, and overrides word stress. Note the final stress in the compound *bāz-kón* ‘opener’ and initial stress in the phrase *bāz kon* ‘open!’, and the stress pattern in the verb phrase; e.g., *raft-é ast* ‘has gone’, but *mí-raft* ‘was going’, *mí-raft-e ast* ‘(apparently) was going’; *dúr raft* ‘went far’, *dúr mí-rav-ad* ‘goes far’.

Foregrounding stress is usually on demonstratives and numbers; e.g., *ín taraf* ‘this way’ and *dó sāt* ‘two hours’. Similarly, discourse insertives, such as *yāʔni* ‘meaning, I mean’, *mónteḥā* ‘in short’, *āyā* ‘I wonder whether’ tend to have initial stress.

Vocative stress is phrase initial; e.g., *hásan* ‘Hasan!’ *áqā-y-e milādī* ‘Mr. M.!’; but may also have final stress according to its function in the discourse. Note: Initial vocative stress is inherited from Indo-European, but is also isoglottic with Semitic and Turkish.

Finally, stress on negation overrides phrase stress; e.g., *dur ná-raft-e ast* ‘has not gone far’.

Always unstressed are: 1) the indefinite marker *-i*; e.g., *kār-i dār-am* ‘I have a matter to discuss’, *soʔāl-e xúb-i bud* ‘that was a good question’; 2) the referential object clitic *-rā* and prepositions; e.g., *kār-rā anjām dād-am* ‘I finished the job’, *az ān-jā* ‘from there’; 3) personal suffixes and endings; e.g., *pedār-am* ‘my father’, *ù-rā dúst mi-dār-am* ‘I love him/her’; 4) phrase clitics; e.g., *mán ham* ‘me, too’, *mán ham ke digar vaxt ná-dār-am* ‘I, too, though, do not have time anymore’.

33.6. Morphophonemic alternation

33.6.1. Verb stems

The Persian verb system is based on present and past stems. The past stem is derived from the present stem by a dental suffix. The “regular” suffix is *-id*, added to the present stem; e.g., *pors* > *pors-íd* ‘ask’, *talab* > *talab-íd* ‘request, search’.

The “irregular” verbs show morphophonemic alternation in the past stem, essentially conditioned changes in the predental consonant. The larger number of those irregular verbs can be synchronically subgrouped accordingly. The following examples cite present and past stems of groups with more than 5 members:

- 1) *b* > *ft*; e.g., *yāb* > *yāft* ‘find’; *kub* > *kuft* ‘pound’.
- 2) *z* > *xt*; e.g., *sāz* > *sāxt* ‘build’; *suz* > *suxt* ‘burn’; *goríz* > *goríxt* ‘flee’; *paz* > *poxt* ‘cook’.
- 3) *h* > *st*; e.g., *jah* > *jast* ‘jump’; *šu* > *šost* ‘wash’; *xāh* > *xāst* ‘want’; but *deh* > *dād* ‘give’.
- 4) *ar* > *ord*; e.g., *bar* > *bord* ‘bear, carry’; *āvár* > *āvórd/āvárd* ‘bring’; but *gozar* > *gozašt* ‘pass’.
- 5) *ār* > *ord*; e.g., *šomār* > *šomórd* ‘count’; *fešār* > *fešórd* ‘pressure’.
- 6) *ār* > *āšt*; e.g., *dār* > *dāšt* ‘hold, have’; *gozār* > *gozāšt* ‘put, let’.
- 7) *ā* > *ud*; e.g., *farmā* > *farmúd* ‘command’; *afzā* > *afzúd* ‘increase’.

The most frequent verbs are also the least regular. They include the following:

Three small subgroups:

8) *oft* > *oftād* ‘fall down’; *ist* > *istād* ‘stand up’; *ferést* > *ferestād* ‘send’.

9) *dān* > *dānést* ‘know’; *tavān* > *tavānést* ‘can’; *gerí* > *geríst* ‘weep’.

10) *čín* > *číd* ‘pile, arrange’; *nešín* > *nešást* ‘sit down’.

Largely irregular verbs include:

11) *kon* > *kard* ‘do, make’; *zan* > *zad* ‘hit’; *band* > *bast* ‘bind, tie’.

šav > *šod* ‘become’; *rav* > *raft* ‘leave’; *gu* > *goft* ‘say’.

xor > *xord* ‘eat, take to self’; *gir* > *geréft* ‘take’.

šenās > *šenāxt* ‘know, acquainted with’; *nevís* > *nevéšt* ‘write’.

12) Suppletive: *ā* > *āmád* ‘come’; *bin* > *did* ‘see’; *bāš* > *bud* ‘be’.

33.7. Old Iranian morphophonemics

Synchronically, it is impossible to determine the conditioning factors for these changes even for the larger groups. Thus, a large subset of the highly complex Old Iranian morphophonology continues to retain morphophonemic function in the verb system of Modern Persian. Diachronically, the past stems of the irregular verbs originate in the Old Iranian perfect participle marked by *-tá*, which was added to the zero grade of the root. The present stems originate in various Old Iranian present stem formations.

For example, the pair *bar* > *bord* ‘bear, carry’ originates in Old Iranian *bár-a* vs. *br-tá*. *bár-a* is the full grade of the root *bar-* with the “thematic” present formant *-a*. *br-tá* is the zero grade of the root with the marker of the perfect participle *-tá*. To give another example, the two stems of the most frequent irregular pair *kon* > *kard* have the following origin: *kon-* derives from the zero grade of the root *kṛ-* with the nasal present formant *-náu*, thus *kṛnáu*. *kard* derives also from the zero grade *kṛ-* with the participial *-tá*, thus *kṛtá*. Similarly, the opposition of *z/x* in the group *sāz* > *sāxt* originates in conditioned variants of **k*, as follows: *z* in *sāz*: **sāk-a* > *sāč-a* before earlier front vowel; *sāč-a* > intervocalic *sāja* in late Old Iranian; *sāj* > *sāz* in Persian. *x* in *sāxt*: **sāk-tá* > fricative *sāxtá* before consonant in Old Iranian. In the case of *sāxtá*, the long grade is the result of analogy to the present stem.

33.8. Suffix *-e* and *-eg-*

The most productive morphophonemic alternation in Persian is the correspondence between final *-e* and *-eg-* before stressed suffixes; e.g., *hafte* ‘week’ and *haftegí* ‘weekly’, *do-sālē* ‘2-year-old’ and *do-sālegí* ‘age of two

years'; *bi-čāré* 'helpless' and *bi-čāregān* 'the helpless'. Related, but irregular and rare, is the suffix *-ej-āt*, with the Arabic suffix *-āt*, as in *ruznāmé* 'news-paper' > *ruznāmejāt* 'the press', *sabzī* 'vegetable' > *sabzijāt* 'vegetable(s)'.

Note: Diachronically, *j* originates in the reflex of Middle and late Middle Iranian loans in Arabic, also found in Arabic *barnāmaj* 'program', Middle Persian *barnāmag*, Modern Persian *barnāmé*. Thus, forms like *ruznāmejāt* and *sabzijāt* are hybrid Arabicized collectives derived from archaized **rōz-nām-ag* and *sabz-īg*.

Diachronically, then, the derivative ending *-é* in Persian derives from *-a* < *-ag*. Such *-a* is still underlying in colloquial Persian, where regular forms like *xunām* 'my house', with a stressed drawl, and *xunā-ro* 'the house (direct object)' evidence the retention of *-a* before the personal suffixes and *-ro*, as opposed to *xuné* elsewhere. Synchronically, therefore, it is possible to posit the rule that prepausal *-ā* changes to *é*, *Cá#* > *Cé#*. Afghan and Tajiki Persian, in fact, retain *-ā* (see below).

33.9. Dialect variation

The major difference between the phonologies of Iranian, Afghan, and Tajiki Persian is found in their vocalic system. By triangulation it becomes evident that Iranian Persian is close to Afghan Persian, but shows the merger of the mid long vowels *ē* and *ō* with the high vowels *i* and *u*. By contrast, Tajiki Persian shows the merger of short mid *e* and long high *i* into *i*, and of short mid *o* and long high *u* into *u*. Systemically, a common underlying vocalic system can be established, which is identical with the system reconstructed for Early Modern Persian:

Tajiki		i		e		u		ú	a	o
Early NP	i	ī	ē	ū	ū	ō	ā	ā		
Afghani	e	i	ē	o	u	ō	a	ā		
Iranian	e	i		o	u		a	ā		

Tajiki *i*, *u*, *a* are short, and unstable; *e*, *ú*, *o* are long, and stable, mid front, central, back vowels, respectively.

Examples are: Iranian *mí-ravam*, Afghan *mē-ravom*, Tajiki *mé-ravam* 'I go'; Iranian *dust*, Afghan *dōst*, Tajiki *dūst* 'friend'; Iranian *ketāb*, Afghan *ketāb*, Tajiki *kitób* 'book'; Iranian *xāné*, Afghan *xāná*, Tajiki *xoná* 'house'.

Bibliography

- Efimov, Valentin A., Vera S. Rastorgueva, and Elena N. Sharova. 1982. "Persidskij, Tadžikskij, Dari." In *Osnovy iranskogo jazykoznanija*, vol. 3, *Novoiranskije jazyki: Zapadnaja gruppa, Prikaspijskie jazyki*, ed. V. Rastorgueva, pp. 5–230. Moscow: Nauka.
- Gaprindashvili, Shota G., and Dzhamshid S. Giunashvili. 1964. *Fonetika persidskogo jazyka*. Tbilisi: „Mecnierba“, Institut Vostokovedenija AN Gruz. SSR.
- Jazayery, Mohammad Ali, and Herbert H. Paper. 1961. "A Reference Grammar of Modern Persian." ms. Ann Arbor: University of Michigan, Department of Near Eastern Studies.
- Lazard, Gilbert. 1957. *Grammaire du persan contemporain*. Paris: Klincksieck. = *A Grammar of Contemporary Persian*, trans. Shirley A. Lyon. Costa Mesa, Calif.: Mazda, 1992.
- . 1963. *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.
- . 1989. "Le persan." In Schmitt 1989: 263–93.
- Nye, Gertrude E. 1955. "The Phonemes and Morphemes of Modern Persian: A Descriptive Study." Ph.D. dissertation, University of Michigan, Ann Arbor.
- Pisowicz, Andrzej. 1985. *Origins of the New and Middle Persian Phonological Systems* (Uniwersytet Jagielloński, Rozprawy Habilitacyjne 101). Krakow: Nakładem Uniwersytetu Jagiellońskiego.
- Poljakov, Konstantin I. 1988. *Persidskaja fonetika: Opyt sistemnogo issledovanija*. Moscow: Nauka.
- Sokolova, Valentina S. 1953. *Očerki po fonetike iranskikh jazykov*. 2 vols. Moscow: Nauka.
- Sokolova, Valentina S., R. L. Memenova, Jurij I. Bogorad, V. A. Livchits, and A. I. Farkhadjan. 1952. *Novye svedenija po fonetike iranskikh jazykov* (Trudy Instituta Jazykoznanija AN SSR). Moscow: Nauka.
- Schmitt, Rüdiger, ed. 1989. *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert.
- Towhidi, Jalil. 1974. *Studies in the Phonetics and Phonology of Modern Persian*. Hamburg: Buske.
- Windfuhr, Gernot L. 1979. *Persian Grammar: State and History of Research* (Trends in Linguistics: State-of-the-Art Reports 12). The Hague: Mouton.
- . 1987. "Persian." in *The World's Major Languages*, ed. Bernard Comrie, pp. 523–46. London: Croom Helm. Repr. in *The Major Lan-*

guages of South Asia, the Middle East and Africa, ed. Bernard Comrie, pp. 108–31. London: Croom Helm, 1990.

———. 1989a. “New Iranian Languages: Overview.” In Schmitt 1989: 246–50.

———. 1989b. “New West Iranian.” In Schmitt 1989: 251–62.

———. 1989c. “Western Iranian Dialects.” In Schmitt 1989: 294–95.

This page intentionally left blank.

Kurdish Phonology

Ernest N. McCarus

University of Michigan, Ann Arbor

Kurdish is an Indo-European language, like Persian a member of the north-west Iranian subgroup. It is divided into two major non-mutually intelligible subdivisions: northern Kurdish, also called *Kurmanji*, found in Turkey, Armenia, and Syria; and southern Kurdish, called *Sorani* in Iraq and *Mukri* in Iran; its speakers number between fifteen and twenty-five million. Kurmanji was suppressed in Turkey until the Gulf War of 1991; there is currently a Kurdish language newspaper published in Istanbul using Latin script. The Kurdish language has flourished in Soviet Armenia, being written at various times in modified Armenian, Cyrillic, or Latin script. In Syria, the Bedir Khan family developed a Kurmanji script based on Latin, but there has not been any appreciable amount of such publications in Syria. In Iraq, Kurdish-language newspapers and magazines have been published and Kurdish language textbooks have been used in Kurdish areas since World War I; since the 1957 revolution, however, there has been an explosion of publications in Kurdish—journals, newspapers, fiction, textbooks, etc. In Iran, Kurmanji has not been supported in written form, although there are some publications in southern Kurdish.

34.1. Phonemes

The description offered here is that of Sulaimania, Iraq, the most prestigious form of Sorani Kurdish and the basis of Modern Standard Kurdish (MSK), the written literary language found in Iraq and Iran today.¹ Kurdish has been actively published in Iraq at least since World War I; MSK was developed by Iraqi Kurdish scholars in the 1970s to serve as a unified written Kurdish language. Its script is based on a modified form of the Perso-Arabic alphabet, and uses diacritics to set up a phonemic alphabet with a

1. The following transcription conventions are used in this chapter: *italics*, morphophonemic; //, phonemic; [], phonetic; ., syllable division; (), optional segment; ~, alternates with; =, becomes; -, morpheme boundary, bound form; /, morpheme alternant; ..., discontinuous morpheme.

Table 34-1. Consonant Phonemes

		Bilabial	Labiodental	Dental	Alveolar pharyngealized	Alveopalatal	Velar	Uvular	Pharyngeal	Glottal
Stops	vl	p		t			k	q		ʔ
	vd	b		d			g			
Affricates	vl					č				
	vd					ǰ				
Fricatives	vl		f	s	ʂ	ʃ	x		ħ	h
	vd		v	z		ʒ	ɣ		ʕ	
Nasals	vd	m		n			ŋ			
Laterals:										
clear	vd			l						
velarized	vd			ɭ						
Vibrants:										
flap	vd			r						
trill	vd			ʀ						
Glides	vd	w				y				

one-to-one correspondence between grapheme and phoneme—except for the vowel *ɪ*, which is not represented at all.

34.1.1. Consonants

Kurdish² has thirty-one consonant phonemes, as listed in Table 34-1. Of these, *ʂ* and *v* are low-frequency phonemes (see § 34.1.1.1.3).

34.1.1.1. Articulation types

34.1.1.1.1. STOPS

The stops are *p b, t d, k g, q, ʔ*. *p t k* are lightly aspirated, whereas uvular *q* and glottal *ʔ* are unaspirated.³ *k* and *g* are heavily palatalized (symbolized

2. Henceforth “Kurdish” refers to the dialect of Sulaimania, Iraq, unless otherwise specified.

3. Kurmanji dialects show phonemic aspiration in the voiceless stops: see Bakaev 1957, Rizgar 1993. This may have been borrowed from Armenian: Kahn 1976.

here by [ʔ] before the high or mid front vowels *i e* and the rounded high front allophone [u] of the phoneme *w*; they are sometimes indistinguishable to the non-native speaker from the affricates *č* and *j*. The contrasts are illustrated by *ke* [kʰe] ‘who?’, *kay* [kʰeɣ] ‘when?’, *kwe* [kʰɛ] ‘where?’, and *gwe* [gʰɛ] ‘ear’.

h tends to be elided when contiguous to a preceding consonant in word compounding, as in *həl* + *hatin* = *həlhatin* ~ *həlatin* ‘sunrise’, *řož* ‘sun’ + *həl* = *řožhəl* ~ *řožəl* ‘east’. It is normally elided in the imperative form *bihenə* = *bhenə* = *bénə* ‘bring!’.

The stops are exemplified by the following:

Labial	<i>penj</i>	‘five’	<i>binin</i>	‘to see’
Dental	<i>tir</i>	‘arrow’	<i>dəst</i>	‘hand’
Velar	<i>karkar</i> ⁴	‘workman’	<i>gənj</i>	‘young’
Uvular	<i>qurs</i>	‘heavy’		
Glottal	<i>ʔaza</i>	‘brave’		

q occurs mainly in Arabic and Turkish loans.

ʔ may occur in any position in the word, although it normally occurs in word-initial position: *ʔəʔawset* ‘it swells’, *nəʔmʔəzani* ‘I didn’t know’, *nəʔ* ‘no’. Word-initial *ʔ* is deleted in word-compounding, as in *řož* ‘sun’ + *ʔawa* ‘set, setting’ = *řožawa* ‘sunset’, *gošt* ‘meat’ + *ʔaw* ‘water’ = *goštaw* ‘soup’; or in elision across word boundaries: see § 34.2.2.1.2.

34.1.1.1.2. AFFRICATES

The unit phonemes *č* and *j* consist of an alveolar stop and alveopalatal sibilant, [tʃ] and [dʒ] respectively: *ča* ‘tea’, *jo* ‘barley’.⁵

34.1.1.1.3. FRICATIVES

The fricatives include *f v x ɣ h s z ʃ ʒ*, as in *fəřš* ‘carpet’, is heavily palatalized in some Kurdish dialects, as in Rowanduzi *fenık* [fʰe:nık] ‘cool’. *v* is rare, although it does occur in some high-frequency words like *həvdə* ~ *həvvə* ‘seventeen’; it is found for the most part in onomatopoeic items, such as *vıřəvıř* (the whirring noise made by a machine) or in loans from northern Kurdish dialects which have a regular phoneme *v* cognate with Sorani *w*, as in *mırov* ‘man’ and *govar* ‘magazine’.

4. The general rule for stress placement in Kurdish is on the last vowel of the word as in *əzanım* ‘I know’. In other cases, stress is marked with an acute accent /' /, as in *bə'lam* ‘but’ and *názanım* ‘I don’t know’, where the stress falls on the first syllable of each word.

5. A rather widespread dialectal variant of this, roughly from Kerkuk to Rowanduz, is respectively [ts] and [dz]: [tsa:] ‘tea’ and [dzo:] ‘barley’.

The sibilants are dental *s* as in *sinə* ‘tray’ and *z* as in *zin* ‘saddle’, and alveopalatal *ʃ* as in *šin* ‘blue’ and *ž* as in *žin* ‘life’. There is also a pharyngealized alveolar sibilant *ʂ* as in *ʂad* ‘hundred’ which, like its Arabic counterpart, requires a back allophone of adjacent vowels, e.g. [ʂad] ‘hundred’ and *ʂaʎ* [ʂaːʎ] ‘year’; the pharyngealized *ʂ* does not occur in many different lexical items, but they are high-frequency ones.

The velar fricatives are voiceless *x*, which is a high-frequency phoneme occurring in native Iranian roots such as *xɪwa* ‘God’ and *xəw* ‘sleep, dream’; and *ɣ*, which appears mainly in words of Arabic origin, such as *ɣəʎət* ‘mistake’. There tends to be confusion between them even in words of non-Arabic origin, as in *bax* ~ *baɣ* ‘garden’ and *xərə* ~ *ɣərə* ‘stud farm’.

The pharyngeal fricatives, voiceless *ħ* and voiced *ʕ*, are also borrowed from Arabic but have spread into Iranian roots as well: *ħəwt* ‘seven’, *ħəftə* ‘week’ (which are of Iranian origin), and *ħəwz* ‘pool’, *ħɪzb* ‘(political) party’ (of Arabic origin). The voiced pharyngeal *ʕ*, as in *ʕajɪz* ‘unhappy’, is often confused with glottal stop *ʔ*, as in Arabic *ʔərz* ~ *ʕərz* ‘earth’ or Iranian *ʔasman* ~ *ʕasman* ‘sky’.

h is found in both syllable-initial position and syllable-final position: *ħərme* ‘pear’, *gunah* ‘sin’.

34.1.1.1.4. NASALS

In addition to bilabial and dental nasals *m* and *n*, there is a velar nasal *ŋ*, which may occur word-finally or intervocalically: *nan* ‘bread’, *man* ‘to remain’, *maŋ* ‘month’, *dəŋɪm* ‘my voice’.

34.1.1.1.5. LATERALS

Sorani Kurdish distinguishes between a dental lateral *l* and a velarized alveolar lateral *ɭ*, as in *gul* ‘leprosy’ and *ɣul* ‘flowers’. *ɭ* does not occur in word-initial position. (In women’s speech in Sulaimania and in some northern Sorani dialects, velar *ɭ* is replaced by *ɾ*, as in *ʔərem* ‘I say’ instead of *ʔəɭem*.)

34.1.1.1.6. FLAPS AND TRILLS

Kurdish shows a contrast between the apical flap *ɾ*, as in *kər* ‘donkey’, and the apical trill *ʀ*, as in *kəʀ* ‘deaf’. They contrast intervocalically in *kəʀə* ‘he is a donkey’ and *kəʀə* ‘he is deaf’. A sequence of two flaps is equivalent to a trill, as in *ʔənerret* ‘it will be sent’. In word-initial position, only the trill occurs.

34.1.1.1.7. GLIDES

The palatal glide *y* and the rounded labiovelar glide *w* are illustrated in *yari* ‘game, play’ and *wa* ‘thus, like this’. The glide *w* has a high front rounded allophone [ɥ] before the high front vowel *i*, as in *naw* [na:w] ‘name’ but *nāwi* [na:ɥi] ‘his name’; or before the mid front *e*, as in *pɾɪyāwek* [pɾya:ɥek] ‘a man’.

34.1.1.1.8. GEMINATION

All consonants except glottal stop and the affricates occur geminate. Gemination may occur either inherently in the word, as in *bərəlla* ‘loose’, *šəmmə* ‘Saturday’, or may result from morphological processes, as in *ʔəneret* ‘he sends’ and *ʔənerret* ‘he is sent’; *ʔənasım* ‘I know’ and *ʔətannasım* ‘I know you (pl.) (*tan*)’. Geminates in Arabic loans may be retained as such but are often simplified, as with *mırəbbə* ~ *mırəbə* ‘fruit jam’.

34.1.1.2. Arabic loans

Iraqi Kurds are bilingual in Kurdish and Arabic; for them their learned language is Arabic, the ordinary source of their technical terms borrowed into Kurdish, although since World War II Kurdish scholars have tried to purify their language of Arabic loanwords. In speech, the Arabic element is still prominent. Modern Standard Arabic and Iraqi Arabic both contain non-Kurdish consonants: the interdental fricatives *θ* and *ð* and the “emphatic” or pharyngealized consonants *t̤* *d̤* *ð̤*. In the Iraqi dialect of Arabic, *d̤* and *ð̤* fuse into a single consonant, both becoming *ð̤*. On more formal occasions, Kurds will pronounce these three Iraqi Arabic consonants as in Arabic; more often, however, they are assimilated to Kurdish consonants, the interdental fricatives becoming sibilants *s* *z* and the emphatics losing their pharyngeal features, *t* *z* *z*. Thus:

<i>θəwrə</i> ~ <i>səwrə</i>	‘revolution’
<i>ðəm</i> ~ <i>zəm</i>	‘blame’
<i>ðulm</i> ~ <i>zulm</i>	‘oppression’
<i>ʔəyyarə</i> ~ <i>təyarə</i>	‘airplane’
<i>qərð</i> , <i>qərð̤</i> ~ <i>qərz</i>	‘loan’

Arabic *ʃ* is generally preserved as such in Kurdish; indeed many native high-frequency words are pronounced with pharyngealized *ʃ*, as in *ʃəg* ‘dog’, *ʃəwz* ‘green’, *ʃəy* ‘safe, healthy, honest’.

The Arabic pharyngeals *ħ* and *ʕ* are high-frequency items in Kurdish, not only being retained in Arabic loans but also replacing *h* and *ʔ* respectively in Iranian words: *ħəwt* ‘seven’, *ħuštır* ‘camel’ and *ʕast* (*ʔast*) ‘place, position’, *ʔasman* ~ *ʕasman* ‘sky’.

34.1.2. Vowels

Vowels are distinguished by height (high/mid/low and close/open), tongue position (front/central/back), and length (short/long); see Table 34-2. There are three short vowels, *ɪ* *ʊ* *ə*, and five long vowels, *i* *u* *e* *o* *a*; long vowels are twice as long as the short ones.

Table 34-2. Vowel Phonemes

		Front	Central	Back	
		Unrounded	Unrounded	Unrounded	Rounded
High	close	i			u
	open			ɪ	ʊ
Mid	close	e			o
	open		ə		
Low			a		

34.1.2.1. Vowel allophones

The unrounded high back vowel *ɪ* becomes rounded when adjacent to the rounded glide *w*, in which case it is equal to *ʊ*: *wítim* ~ *wótim* ['wutɪm] 'I said', *jíwan* [jʊ'wa:n] 'beautiful'. *ɪ* never occurs in word-final position (but see § 34.2.2.2.2). (See § 34.2.2.1.1 for syncope of *ɪ*.)

The vowel *ə* has the principal allophonic type open mid central vowel [ə], as in *gǝrmə* ['gərmə] 'it is hot', *ʔəwəta* [ʔəwə'ta:] 'there (it) is'; it has the allophone open mid front vowel [ɛ] when it precedes the glide *y* in monosyllables, as in *kəy* [kɛy] 'when?', or when it is followed by *y* in the following syllable: *hǝyə* ['hɛyə] 'there is, there are'; it remains shwa before *y* in a non-word-final syllable, as in *ʔəybinim* [ʔəybi:'nim] 'I see him'.⁶

34.1.2.1.1. LENGTH

All long vowels, *i u e o a*, are shortened when occurring unstressed word-finally: *čáki* [tʃa:'k'i:] 'goodness, good (n.)' but *čáki* [tʃa:k'i] 'you are well'. They are also shortened when occurring before a consonant cluster, as in *ʔənum* [ʔə'nu:m] 'I sleep' but *nústim* ['nustɪm] 'I slept'.

The long high vowels are shortened before homorganic glides: *číyə* [tʃiyə] 'What is it?', *zúwə* [zuwə] 'It is early.'

Otherwise, long vowels are always long, as in *ʔasman* [ʔa:s'ma:n] 'sky', *bǝlam* [bəla:m] 'but'.

34.1.2.1.2. VOWEL DISTRIBUTION

Vowels do not occur in syllable-initial position. The phoneme glottal stop *ʔ* occurs before an otherwise word-initial vowel, as in *ʔaw* 'water', *ʔewə* 'you (pl.)', *ʔəzani* 'you know'.

6. The Kirkuk dialect of Arabic shows also the rounded high close front vowel *ü*, which, together with *u*, corresponds to Sulaimania Kurdish *u*; for example, Kirkuki *dür* corresponds to Sul. *dur* 'far, distant' (see Ahmad 1986).

The short high vowels do not occur word-finally, except for *ɪ* as the result of dental stop deletion; see § 34.2.2.2.

Illustrations of vowels:

<i>i</i>	<i>ʔiʃ</i> ‘work’, <i>ʒin</i> ‘life’, <i>si</i> ‘thirty’
<i>u</i>	<i>ʔutel</i> ‘hotel’, <i>kuʃ</i> ‘hunchback’, <i>xanu</i> ‘house’
<i>ɪ</i>	<i>kirdim</i> ‘I did, made’, <i>dɪl</i> ‘heart’, <i>nízɪm</i> ‘low’
<i>ʊ</i>	<i>ʔustaz</i> ‘professor’, <i>kuʃ</i> ‘son’, <i>qutabi</i> ‘student’
<i>e</i>	<i>ʔewə</i> ‘you (pl.)’, <i>ʃew</i> ‘evening meal’, <i>ʔəçe</i> ‘goes’, <i>se</i> ‘three’
<i>ə</i>	<i>ʔəwan</i> ‘they’, <i>ʃəw</i> ‘evening’, <i>tələbə</i> ‘student’, <i>gəwrə</i> ‘big’
<i>o</i>	<i>ʔosta</i> ‘master (craftsman)’, <i>koʃ</i> ‘gathering (of people)’, <i>bo to</i> ‘for you (sg.)’
<i>a</i>	<i>ʔasman</i> ‘sky’, <i>ʔaza</i> ‘brave, free’, <i>ʔəwəta</i> ‘there is ...’

34.2. Morphophonology

34.2.1. Allomorphy

Certain lexemes show phonological changes that follow phonological rules different from those given below in § 34.2.2 on morphophonemics; many of these items involve shwa ə. Allomorphic rules generally preserve morphemic identity, since the morphophonemic rules would for the most part trigger shwa deletion and destroy the identity of the morpheme.

The suffixed members of the discontinuous (envelope) morphemes *ʔəm...ə* ‘this’ and *ʔəw...ə* ‘that’ occur after consonants; glides are inserted when following vowels. After back vowels *u o*, the homorganic glide *w* is inserted, whereas *y* is inserted after all other vowels, thus:

<i>dəst</i>	‘hand’	<i>ʔəmdəstə</i>	‘this hand’
<i>mu</i>	‘hair’	<i>ʔəwmuwə</i>	‘that hair’
<i>wɪʃə</i>	‘word’	<i>ʔəmwɪʃəyə</i>	‘this word’
<i>gorani</i>	‘song’	<i>ʔəwgoraniyə</i>	‘that song’
<i>dərga</i>	‘door’	<i>ʔəmdərgayə</i>	‘this door’

The copula suffix *-ə* ‘is’, which is unstressed, has the unstressed allomorph *-eti* after the morpheme *-y* (3sg.), as in *piyaw* ‘man’ + *-y* ‘his’ + *-ə* ‘is’ = *piyáwyeti* ‘he is his (hired) man’. In all other instances of *-ə* as copula, the allomorph *-wə* occurs after *u*, as in *zu* ‘early’, *zúwə* ‘it is early’, and the allomorph *-yə* occurs after all other vowels, as in *gorani* ‘song’, *goraníyə* ‘it is a song’; *diro* ‘lie’, *díróyə* ‘it is a lie’; and *dərga* ‘door’, *dərgáyə* ‘it is a door’.

The vocative suffixes *-ə* (sg.) and *-inə* (pl.) occur on nouns: *-ə* occurs after consonants on monosyllables, as in *kúʃə* ‘boy!’ but has the allomorph *-wə* after nouns in *-o*, as in *hérowə* ‘Hero!’ (girl’s name). *-inə* occurs on all forms,

as in *xəʃqinə* '(You) people!' (*xəʃq*). The vocative suffixes are unstressed, as all vocatives bear word-initial stress (see § 34.4.1.3).

The prepositions *bə* 'with; to' and *lə* 'in; from' take the forms *pe* and *le* when followed by a suffixed pronoun object or by no object at all: *bə to ʔəʃem* ~ *pet ʔəʃem* 'I say to you (*to*, *-t*)'; *mərḥəbə lə to ʔədəm* 'I give you my hello's = I greet you' and *mərḥəbəledan* 'to greet'; *ḥəzi le ʔkəm* 'I like it (*-i*)'.

The indefinite suffix, which is unstressed, has the allomorphs *-ek* after consonants and *-yek* after vowels: *maʃ* 'house', *maʃlek* 'a house'; *dərğa* 'door', *dərğáyek* 'a door'; *ḥərme* 'pear', *ḥərmeýek* 'a pear'; *dıróyek* 'a lie'. The *k* of this suffix is often dropped: *piyáwek* ~ *piyáwe* 'a man'.

The verbal suffix *-əwə* 'back, again', which is unstressed, has the allomorph *-rəwə* when added to a verb stem ending in a vowel, as in *ʔəčit* 'you (sg.) go' and *ʔəčítəwə* 'you go back, you return' (stem = *-č-*), but *bíkə- + -rəwə* = *bíkərəwə* 'open it!' (stem = *-kə-*).

The noun suffix *-əwə* (see § 34.4.1.5) is monoallomorphic, occurring after both consonants, as in *lə ʔutéləwə* 'from the hotel', and vowels, as in *lə bəyáwə* 'from Baghdad (*bəya*)' (see § 34.2.2.1.2 for elision of shwa).

The verbal suffix *-e* 'to' (indirect object) may only be suffixed on the verb *dan* 'to give', present stem *-də-*; if it follows a vowel, the allomorph *-re* occurs: *ʔətdəme* 'I give it to you' but *bímdəre* 'give it to me!'.

34.2.2. Morphophonemics

Morphemes with morpheme alternants are listed in § 34.2.1 on allomorphy; the present section presents the rules governing changes resulting from morphological and phonological processes in the language at large.

34.2.2.1. Vowels

34.2.2.1.1. SYNCOPE OF *ɪ*

The vowel *ɪ* is elided when it occurs unstressed between two single consonants if the second of these two consonants is followed by a vowel: *kiteb* 'book' is phonemically /kteb/, with automatic syncope of *ɪ*. Additional illustrations: *nízim* 'low' : *nízimə* = /nízɪmə/ 'it is low'; *piyaw* = /pyaw/ 'man', *xíwa* = /xwa/ 'God'. If the vowel is stressed, syncope is blocked: *bíra* 'brother' is phonemically /bra/; in the vocative form, however, where the first vowel of the word is stressed, syncope of *ɪ* is blocked: *bíra* = phonemic /bíra/ 'O brother!'.

34.2.2.1.2. ELISION

There are no vowel sequences in Kurdish (but see Dental stop elision, § 34.2.2.2.2). The following changes take place when vowels come into contact with each other as the result of morphological processes.

The high back unrounded vowel *ɪ* elides when it comes in contact with any vowel:

qəɭəməkə ‘the pen’ + *-it* (2sg.) = *qəɭəməkət* ‘your pen’

bɪxo ‘eat!’; *bɪ-* (imperative marker) + *-i-* ‘it’ + *-xo-* ‘eat’ = *bɪxo* ‘eat it!’

Two identical long high or mid front vowels undergo elision of one vowel:

i + *i* = *i* *bini-* ‘saw’ + *-in* ‘we’ = *binin* ‘we saw’

e + *e* = *e* *ɭəɭe-* ‘say’ + *-et* (3sg.) = *ɭəɭet* ‘says’

An exception to this is the indefinite suffix *-ek*; see under § 34.2.1.

The short vowel *ə* followed by another *ə* elides:

təxtə ‘board’ + *-əkə* (definite article) = *təxtəkə* ‘the board’

ə adjacent to a long low vowel undergoes elision:

ɭəkə- ‘do’ + *-at* (3sg.) = *ɭəkat* ‘does’

When a monosyllable ending in *ə* is followed by a deictic or pronoun beginning with *ɭ*, both *ə* and *ɭ* are elided:

bə ‘with’ + *ɭəmɔrə* ‘this way’ = *bəmɔrə* ‘in this way’

lə ‘in’ + *ɭerə* ‘here’ = *lerə* ‘here, in this place’

bə ‘to’ + *ɭəwan* ‘they’ = *bəwan* ‘to them’

lə ‘from’ + *ɭəmə* ‘this’ + *u* ‘and’ + *paš* ‘after’ = *ləməwpaš* ‘from now on’
(for *u* > *w* see § 34.2.2.1.4).

34.2.2.1.3. GLIDE INSERTION

Glide insertion is triggered by sequences of a long vowel and a long low vowel or a stressed long high vowel.

A sequence of *-u* followed by *a* or by a stressed long high vowel undergoes *w*-insertion:

bə ‘in’ + *ɭəm...ə* ‘this’ + *zu* ‘early’ + *-an* (pl.) = *bəmzuwanə* ‘in the next few days, soon’

bə ‘in; with’ + *zu* ‘quick’ + *-i* (abstract noun) = *bə zuwɪ* ‘quickly’

All other sequences of long vowel followed by a long vowel or stressed high vowel undergo *y*-insertion:

bɪra ‘brother’ + *-an* (pl.) = *bɪrayan* ‘brothers’

ɭəw...ə ‘that’ + *hərme* ‘pear’ + *-an* (pl.) = *ɭəwhərmeyanə* ‘those pears’

ɭəm...ə ‘this’ + *dɪro* ‘lie’ + *-an* (pl.) = *ɭəmdɪroyanə* ‘these lies’

ɭaza ‘free’ + *-i* (abstract noun) = *ɭazayɪ* ‘freedom’

After long vowels, unstressed *i* *u* change to glides: see § 34.2.2.1.4.

34.2.2.1.4. CHANGE

Unstressed long high vowels change to homorganic glides after a stressed (syllabic) vowel:

bu- ‘was’ + *-in* ‘we’ = *buyn* ‘we were’; cf. *řóyštin* ‘we went’

mežu ‘history’ + *-u* ‘and’ + *čirok* ‘story’ = *mežuw čirok* ‘history and tales’

řamadə ‘ready’ + *-it* ‘you are’ = *řamadəyt* ‘you are ready’; cf. *čónit* ‘how (čon) are you?’

-u ‘and’, *jɪwan* ‘beautiful’, *pəti* ‘pure’: *jɪwánu pəti* ‘beautiful and pure’
vs. *pətiw jɪwan* ‘pure and beautiful’

The back vowels change to *w* before a following stressed long vowel:

řenum ‘I sleep’ (stem = *-nu-*): *řenwet* ‘he sleeps’

řəřom ‘I go’ (stem = *-řo-*): *řəřwat* ‘he goes’

34.2.2.1.5. FUSION

In polysyllabic words, front and back long high vowels normally fuse with *ə* to yield respectively front and back long mid vowels:

i + *ə* = *e* *yari* ‘game, play’ + *-əkə* ‘the’ = *yarekə* ‘the play’

u + *ə* = *o* *řarəzu* ‘desire, wish’ + *-əkə* ‘the’ = *řarəzokə* ‘the desire’

If fusion does not take place, the resultant combination receives special focus, as in *řarəzuwəkə* ‘the *desire*’.

34.2.2.2. Consonantal processes

34.2.2.2.1. CLUSTER REDUCTION

The cluster *nd* may stand as it is, it may be reduced to *n*, or—probably most often—it may be replaced by *ŋ*: *čənd* ‘how much?’, *čəndit řəwe* ~ *čənit řəwe* ~ *čəŋit řəwe* ‘how much do you want?’, *šəməndəfər* ~ *šəməŋəfər* ‘train’.

34.2.2.2.2. DENTAL STOP ELISION

The dental stops undergo optional elision under the following conditions.

The morpheme *-t* of the 2sg. pronoun suffix *-it-* is optionally deleted in word-final position: *dəstɪt* ~ *dəstɪ* ‘your hand’. The vowel *ɪ* occurs in word-final position only after dental stop elision; see below.

The *t* of the 2sg. and 3sg. verb inflections *-it* and *-et* respectively are normally elided if not followed by another suffix: *řəčit* ~ *řəči* ‘you go’ but *řəčítəwə* ‘you go back’; *řəlet* ~ *řəle* ‘he says’ but *řəlétəwə* ‘he repeats’.

d is usually elided as the first segment of the locative suffix *-da*, as in *lə šárda* ~ *lə šára* ‘in the city’; *bəyda* ‘Baghdad’ is normally pronounced *bəya*.

d in postvocalic position may be replaced by the unrounded high back open vowel *ɪ*, especially in proper names: *řəħməd* ~ *řəħməɪ* ‘Ahmad’; *řədəm* ~ *řəɪəm* ‘I give’; *qadir* ~ *qair* ‘Qadir’; *bəd* ~ *bəɪ* ‘bad’. This is the only instance

in which a vowel sequence may occur. It should be noted that this is clearly a sequence of two vowels with no suggestion of an intervening glide. See above for the other instance of the occurrence of *ɪ* in word-final position.

34.2.2.2.3. DENTAL STOP ASSIMILATION

Dental *-t* is optionally assimilated to a following dental nasal in normal rapid speech, as in *bırsıt nıyə? ~ bırsın nıyə?* 'Aren't you hungry?', *hiç zəhmət nıyə ~ hiç zəhmən nıyə* 'It's not at all difficult.'

Word-initial voiced dental stop *d* in *dan* 'to give' is optionally assimilated to a preceding bilabial or dental nasal in normal speech register: *bıryárman da ~ bıryárman na* 'we decided'; *lem dayt ~ lem mayt* 'I hit (past) you.'

34.2.2.2.4. ASSIMILATION AS TO VOICE

ş and *b* undergo assimilation as to voice. *ş* becomes voiced when in full contact with a following voiced consonant, as in *paş* 'after' but *paž niwəro* 'afternoon', and *baş* 'good' but *bážbu* 'it was good'.

b is devoiced when it comes in full contact with a following voiceless consonant, as in *čí bıkəm? ~ čípəkəm?* 'What shall I do?'

34.3. Syllable structure

Following are the permissible syllable structures on the underlying morpho-phonemic level; the period denotes syllable division.

CV	<i>lə</i> 'in, from', <i>tə.lə.bə</i> 'student'
CVC	<i>mın</i> 'I', <i>hər</i> 'only', <i>mək.təb</i> 'school'
CVCC	<i>kurd</i> 'Kurd', <i>dost</i> 'friend', <i>wíst.man</i> 'we wanted' (preterite tense)

Three-consonant clusters result when syllables are combined in word formation: *kírdman* 'we made (it)'. On the phonemic level the syllable pattern /CCVCC/ occurs: /fróšt.man/ 'we sold'. When a word like *xošəwíst* 'beloved, dear' is made superlative with the addition of *-tírín* giving *xošəwísttírín* 'dearest', on the phonemic level *ɪ*-syncope will create a phonemic cluster of four consonants: /xošəwísttrín/.

34.4. Stress placement

34.4.1. Word stress

Stress in Kurdish as a rule falls on the last vowel of the stem of the word; that is, inflections on the whole are not stressed. Derivational suffixes, which create new stems, on the other hand do take word stress: *dırež* 'long' + *i*

(abstract noun) = *dīreži* ‘length’. Stress is not entirely predictable, however, since there are many cases of exceptions to this generalization. In the case of verbs, subject marker suffixes are normally stressed in the present tense but are never stressed in the past tenses, so creating contrasts with nouns. To illustrate: *hatin* with ultima stress is the infinitive ‘to come’; *hátin*, however, with penultimate stress, is the verb ‘they came’, composed of the past stem *hat-* ‘came’ and the 3pl. intransitive ending *-in* ‘they’, which never takes stress. Word-final stress is not marked; stress is marked / ‘/ on all other syllables.

A third category of suffixes is the suffixed pronouns, which are unstressed. It is also significant that these suffixes also serve as subject marker inflections on past tense transitive verbs:

<i>fuʔádīm bini</i>	‘I saw Fuad’	<i>fuʔádmān bini</i>	‘we saw Fuad’
<i>fuʔádīt bini</i>	‘you saw Fuad’	<i>fuʔádtān bini</i>	‘you saw Fuad’
<i>fuʔádi bini</i>	‘he saw Fuad’	<i>fuʔádyān bini</i>	‘we saw Fuad’ ⁷

Following are the most important exceptions to the general rule of word-final vowel stress.

34.4.1.1. *Verbs*

Negative prefixes and the subjunctive/imperative prefix *bí-* receive word stress; past tense subject markers and the copula suffixes do not receive word stress. Illustrations: *ʔaxīwenīm* ‘I read’ but *náxīwenīm* ‘I do not read’, *bíxīwenə* ‘read!’, *xīwéndīm* ‘I read’ (preterite tense), and *nəmxīwend* ‘I did not read’.

The verbal suffix denoting indirect object *-e* ‘to’, which is suffixed only to the verb *dan* ‘to give’, is not stressed: *ʔədəm* ‘I give’ + *-t-* ‘you’ (sg.) + *-e* ‘to’ = *ʔətdəme* ‘I give (it) to you’ (pronouns may be infixed in the present tense between the imperfective aspect prefix *ʔə-* and the stem of the verb, in this case *-də-*).

The verbal suffix *-əwə/-rəwə* ‘back, again’ (see § 34.2.1) is unstressed: *ʔəydəm* ‘I give it (-y-): *ʔəydəməwə* ‘I give it back’; *bíle* ‘say (sg.) it!’ and *bílerəwə* ‘repeat it!’

Negative prefixes on verbs are stressed: *ná-*, found on the present indicative verb; *nə-*, found on the present subjunctive and all past tenses; and *mə-*, occurring with the negative imperative:

Present tense indicative:	<i>ʔədəm</i> ‘I give’	<i>nádəm</i> ‘I do not give’
Present tense subjunctive:	<i>bídəm</i> ‘(if) I give’	<i>nədəm</i> ‘(if) I do not give’

7. Subject markers of past tense transitive verbs are suffixed to an earlier word in the clause.

Imperative:	<i>bîdê</i> 'give!'	<i>mêdê</i> 'don't give!'
Preterite tense indicative:	<i>dam</i> 'I gave'	<i>nêmda</i> 'I didn't give'

Note that the imperative form of the verb takes word-initial stress like the vocatives of nouns (see § 34.4.1.3).

Negative prefixes on nouns and adjectives, by contrast, are not stressed: *xoš* 'good' – *naxoš* 'unpleasant', *nəxoš* 'ill'; *nasraw* 'known, familiar' – *nə-nasraw* 'stranger'.

34.4.1.2. Nouns

The suffixes *-an* (plural) and *-əkə* (definite) are stressed, but all other noun suffixes are unstressed. Illustrations: *kîč* 'girl' and *kîčan* 'girls', *kîčəkə* 'the girl' but *kîček* 'a girl'. The plurals *ga* 'cow' – *gagal* 'cows' and *de* 'village' – *dehat* 'villages' also take word stress.

A few nouns of place and time take the relic unstressed locative suffix *-e*, as in *šāre* 'in the city' and *šəwe* 'in the evening'.

34.4.1.3. Vocatives

Nouns in vocative function are stressed on the first (underlying) vowel in the word. Thus, *mamosta* 'teacher', *mámosta* 'O teacher!', 'Sir!'; *?əhməd* 'Ah-mad!'. In a word like *bíra* 'O brother!', the vowel *ɪ*, being stressed, does not undergo syncope (cf. § 34.2.2.1.1).

Imperative verbs can be construed as vocatives: see § 34.4.1.1.

34.4.1.4. Adjectives and Adverbs

Adjectives may take the same suffixes as nouns, but may in addition, like adverbs, take the comparative suffix *-tír* 'more' and the superlative suffix *-tírin* 'most', which take word stress: *čak* 'good', *čaktír* 'better', *čaktírin* 'best'; *xera* 'fast', *xerátír* 'faster'. These degree suffixes are exceptional in that they are stressed even though they are inflectional suffixes.

34.4.1.5. Particles

Most particles are stressed on the word-initial vowel, like *bəlam* 'but', *čúŋkə* 'because', *dwéne* 'yesterday'. Exceptions are *hətáku* 'until, up to' and the noun-like adverbs *?está* 'now', *?erə* 'here, this place', and *?əwé* 'there, that place'.

The unstressed suffix *-əwə* may occur on nouns with the meaning 'from', as in *lə kərkúkəwə* 'from Kirkuk', as well as on particles, as in *lə ?estáwə* 'from now (*?esta*) on'; it may occur on verbs with the meaning of 'back' or

‘again’, as in *bílen* ‘say (pl.) it!’ and *bílenəwə* ‘say it again, repeat it!’; *ʔəyandəm* ‘I give them’ and *ʔəyandəməwə* ‘I give them back, I return them’.

In the case of infinitives, the suffix *-əwə* assumes the character of a derivational rather than an inflectional suffix and, as part of the stem, does take word-final stress: *kirdín* ‘to do, make’ and *kirdínəwə* ‘to open’.

34.4.2. *Primary and secondary stress*

Both primary and secondary stress are found on the word level. For example, the demonstrative envelope *ʔəm...ə* ‘this’ takes primary stress on the first member and secondary stress on the second one, with the enclosed noun being unstressed or, more properly, receiving tertiary stress, as in *ʔəmšəqamə* ‘this street’, *ʔəmšəqaməjiwanə* ‘this beautiful street’. Primary and secondary stress placement may be reversed for the sake of emphasis.

34.4.3. *Sentence stress*

In phrases and clauses, certain classes of words are normally stressed, whereas others as a rule are not.

34.4.3.1. *Noun Phrases*

The following classes of words receive phrase stress and are followed by unstressed nouns.

1) Interrogatives: *čənd pɪyaw* ‘how many men?’, *kám wišəyə* ‘which word?’

2) Numerals: *də pɪyaw* ‘ten men’, *pənʃ šəqam* ‘five streets’

3) Quantifiers: *həmú pɪyawek* ‘each man’, *čənd pɪyawek* ‘a few men’, *čənd řožek* ‘a few days’. *zor* ‘many’ (adjective), ‘very’ (adverbial) is always stressed: *zór pɪyaw* ‘many men’, *zór čak* ‘very good’.

34.4.3.2. *Ezafe Phrases*

The liaison morpheme *-i-* (“ezafe”) links a noun with a following modifying word, phrase, or clause; when the following modifier is a word or a string of conjoined words, those modifiers are stressed rather than the modified noun: *pɪyawek-i-čák* ‘a good man’, *řožek-i-wá* ‘a day like this’, *šəqamek-i-tazəw dúru jɪwán* ‘a beautiful, long (*dur*), new (*tazə*) street’.

34.4.3.3. *Clauses*

In clauses, nouns, adjectives and adverbs, interrogatives, preverbals, and negative verbs tend to receive sentence stress. Illustrations:

Nouns: *pıyawəkə hat ləgəlma* 'The man came with me.'

Adjectives/Adverbs: *kurdi bāš ʔəzani* 'You know Kurdish well (*baš*).'

Interrogatives: *ʔəw kəy hat bo ʔerə* 'When did he come here?'

Preverbal: *da* 'down' in *kıtebəkāni ləsər təxtəkə də na* 'He put the books down on the board.'

Negative verbs: *kıtebəkān ləsər təxtəkə də ʔanem* 'I place the books down on the board', but *kıtebəkān ləsər təxtəkə da nānem* 'I don't place the books down on the board'; *kıtebəkān ləsər təxtəkə da mənə* 'Don't (*mə-*) put the books down on the board.'

The exhortative particles *bā* 'let's ...' and *də*, and *dəy* (asseveration) normally take sentence stress: *bāšə, bā bıçin* 'Good, let's go!', *də bıro dəy* 'go, go on, get a move on!'

Prepositions do not normally take sentence stress, but the preposition *bō* 'to, for' does as a rule, as in *bā bıçin pəkəwə bō bazar* 'Let's go together to the marketplace.'

The adverb *wa* 'thus, like this' is a special case: In its denotative meaning it is stressed, but when unstressed it assumes the meaning of immediacy of action, as in *wā hatım* 'I came thus, I came like this' but *wa hātım* 'I'm coming, I'm on my way now!'

34.4.4. *Emphasis*

Any word in a sentence can receive special focus or emphasis by receiving primary stress, signaled here by non-italicization and the acute accent:

<i>to kurdi bāš ʔəzanit</i>	'You know Kurdish well.' (neutral)
<i>tó kurdi baš ʔəzanit</i>	' <i>You</i> know Kurdish well.'
<i>to kurdí baš ʔəzanit</i>	'You know <i>Kurdish</i> well.'
<i>to kurdi bāš ʔəzanit</i>	'You know Kurdish <i>well</i> .'
<i>to kurdi baš ʔəzanít</i>	'You <i>know</i> Kurdish well.'

Bibliography

- Abdulla, Jamal Jalal, and Ernest N. McCarus. 1967. *Kurdish Basic Course: Dialect of Sulaimania, Iraq*. Ann Arbor: University of Michigan Press.
- Ahmad, Abdul-Majeed Rashid. 1986. "The Phonemic System of Modern Standard Kurdish." Ph.D. dissertation, University of Michigan.
- Bakaeu, Ch. Kh. 1957. *Kurdsko-Russkiy Slovar*. Moscow: Akademia Nauk.
- Bedir Khan, Emir Djeladet, and Roger Lescot. 1970. *Grammaire Kurde (Dialecte Kurmandji)*. Paris: Maisonneuve.

- Blau, Joyce. 1980. *Manuel de Kurde. Dialecte Sorani*. Paris: Klincksieck.
- Kahn, Margaret. 1976. "Borrowing and Variation in a Phonological Description of Kurdish." Ph.D. dissertation, University of Michigan.
- MacKenzie, D. N. 1961. *Kurdish Dialect Studies*, 2 vols. London: Oxford University Press.
- McCarus, Ernest. 1958. *A Kurdish Grammar: Descriptive Analysis of the Kurdish of Sulaimania, Iraq*. New York: American Council of Learned Societies.
- Rizgar, Baran. 1993. *Kurdish-English English-Kurdish Dictionary*. London: Author. [Kurmanji]
- Wahby, Taufiq, and C. J. Edmonds. 1966. *A Kurdish-English Dictionary*. Oxford: Clarendon. [Sorani]

Ossetic Phonology

David Testen

University of Chicago

35.1. Background

Ossetic¹ is a modern Iranian language spoken in the central Caucasus region. A member of the Northeast Iranian branch, Ossetic is regarded as a direct descendant of ancient Scythian and of medieval Alanic, languages spoken by the nomadic Iranians of the steppes north of the Caucasus and the Black Sea. For neither Scythian nor Alanic have texts providing direct evidence of the nature of the language survived, but Scythian personal names preserved in inscriptions from the Greek Black Sea colonies show close similarities to modern Ossetic forms—cf., e.g., ΦΟΥΡΤΑΣ alongside West Ossetic *furt* ‘son’, ΙΝΣΑΖΑΓΟΣ alongside West Ossetic *insæj*, East Ossetic *ssæz* ‘twenty’ (Zgusta 1955)—as do several of the Alanic phrases quoted by the twelfth-century Greek poet Johannes Tzetzes—cf. the form ταπαρχας, glossed by Tzetzes as ‘good day to you’, which may be compared to the modern Ossetic greeting *dæ bon xorz* ‘your day (be) good’ (Abaev 1949: 254–59). Other extinct languages akin to Scythian, Alanic, and Ossetic are Sogdian and Khwarezmian. Ossetic’s closest relative among the living Iranian languages is Yaghnobi, a descendant of Sogdian surviving in Tajikistan.

Geographically, Ossetia is divided by the Great Caucasus range into North Ossetia and South Ossetia. North Ossetia, with an area of 8,000 km² and a population of 643,000 (1989 est.), is an Autonomous Republic of the Russian Federation lying in the valley of the upper and middle Terek and its tributaries, the Ardon, the Gizeldon, the Urux, and the Fiagdon; the North Ossetian capital is Vladikavkaz (formerly Oržonikizë). South Ossetia, the capital of which is Cxinval (formerly Dzæwžiqæw), is located in the north-central region of the Georgian Republic; it has an area of 3,800 km² and a

1. The author is indebted to Bill J. Darden, Nancy L. Dray, and Elisa Steinberg for their comments on an earlier draft of this paper.

The terms “Ossetic” and “Ossetian” found in the European languages are derived ultimately from Georgian *O(v)s-eti* ‘the territory of the *O(v)s-i*.’ The Ossetic self-designation is *Ir*, the native name for the language is *Iron*, and the territory is known as *Iriston*.

population (1979) of 103,000, of which 65,000 are Ossetians. During the Soviet period, South Ossetia was an Autonomous Oblast' of the Georgian Republic. Since the eclipse of Soviet power, the status of South Ossetia has become a matter of contention; it was dissolved as an autonomous entity by the Georgian legislature in December 1990 amid ongoing ethnic violence.

The primary dialect division of Ossetic lies between the western dialect (Digor), spoken in the western territory of North Ossetia, and the eastern dialect (Iron), spoken in most of North Ossetia and all of South Ossetia. In the Wællagkom region a transitional dialect between Digor and Iron is spoken (Isaev 1966: 101–12). The Ossetic literary language is a form of Iron, and is traced back to the works of Xetagkati K'osta (Russ. Kosta Xetagurov), the Ossetic national poet (1859–1906).

35.2. Script

The earliest known text in Ossetic, a funeral inscription discovered in 1888 in the Zelenčuk River region and dated, on paleographic grounds, to the tenth or eleventh century A.D., used Greek characters (Zgusta 1987). In the eighteenth and nineteenth centuries Orthodox missionaries experimented with creating writing systems for Ossetic: Gaj Takaov published an Ossetic catechism (1798) using a version of the Old Slavonic Cyrillic script, and in the 1820s Ivane Ialʒuzize employed an Ossetic alphabet based on the Georgian *xucuri*. In 1844, the noted linguist A. Sjögren introduced a graphic system based on Russian Cyrillic which was substantially the same as the alphabet in use today. During the Soviet period the Cyrillic script was temporarily replaced by a Latin script (1923–1938) and, in South Ossetia, by a Georgian script (1938–1954).

The modern alphabet consists of forty-three characters, of which eight (ë, ž, š, šč, ʷ (the “soft sign”), è, ju, and ja) are found only in Russian loanwords. An important difference between Ossetic Cyrillic and Russian Cyrillic is the use of the Russian “hard sign” (ə) to form digraphs. In conjunction with the Cyrillic signs for the voiceless obstruents (i.e., pə, tə, cə, čə, kə) this sign indicates the phonemes of the glottalic ejective series (hereafter indicated by an apostrophe—p', t', c', č', k'). A hard sign together with g (gə) marks the ɣ; a hard sign with the sign for the voiceless uvular continuant x (i.e., xə) marks q. In the absence of symbols in Russian for the voiced affricates ʒ, ʒ, these phonemes too are spelled with digraphs (dz, dʒ). The so-called “fourth series” of obstruents, which are pronounced voiceless and unaspirated, are

normally rendered using the characters for the voiceless obstruents, but not infrequently are represented with the voiced characters (*ræxčĭ/ræxžĭ* ‘soon’); likewise, geminate obstruents, despite the fact that they are uniformly pronounced voiceless, may be spelled with doubled voiceless characters (*pp*, etc.), with doubled voiced characters (*bb*), or with sequences of voiced and voiceless characters (*bp*, *pb*). The weak low vowel is rendered using a symbol *æ* introduced into Cyrillic; the Iron high central vowel (hereafter *ĭ*) is spelled using the Russian *ѣ* (*jery*).

35.3. The Phonemic Inventory

See Tables 35-1, 35-2, and 35-3.

35.4. Consonantal Phonemes

35.4.1. Occlusives

The voiceless occlusives *p*, *t*, *c*, *k* and (in Iron) *č* are pronounced with strong aspiration in most positions ([p^h], [t^h], [ts^h], [k^h], [tʃ^h]). Although the aspiration is normally lost after the voiceless continuants (*f*, *x*, *s*), the presence of an intervening morpheme boundary will often cause the aspiration to be retained even in this position, the aspiration thus not infrequently serving to indicate the morphological composition of the word—cf. *stajĭn* ‘grow tired (inf.)’ vs. *s-t^(h)ajĭn* ‘thaw out, melt (inf.)’ (perfective prefix *s-* + *t^hajĭn* ‘thaw’), and *xæst-ĭ* ‘in the war (*xæst* + locative *-ĭ*)’ vs. *xæst-t^(h)-ĭ* ‘in the obligations (*xæst* + plural *-t^h(æ)* + locative *-ĭ*)’. They also lose their aspiration when geminated (*xabær-ttæ*, plural of *xabar* ‘news’), and consequently merge with the voiced occlusives (see § 35.4.1.1).

Early Iranian voiceless stops have generally become voiced outside of initial position; **-t-* and **-k-* remained stops (*d*, *g*) and **-č-* remained an affricate (*č*), while **-p-* became a continuant (*v*)—cf. Iron *xid*, Digor *xed* ‘bridge’ < **haitu-*, Iron *qug*, Digor *yog* ‘cow’ < **gaw-ka-*, Iron *æxsæv*, Digor *æxsævæ* ‘night’ < **xšapan-*.² These historical changes are recapitulated in the compound-formation rules of the modern language: word-initial voiceless obstruents typically shift to their voiced counterparts when they come to be in medial position—cf. Iron *x^higæs* ‘swineherd’ (*x^hĭ* ‘pig’ + the root of *kæs-in* ‘look’), *divænd* ‘vacillating’ (< Iranian **dvi-* ‘two-’ + *fænd* ‘will’), *wæłzarm*

2. Iranian **p* in general has become a continuant: Iron *firt*, Digor *furt* ‘son’ < **puθra-*, Iron *jæfs*, Digor *æfsæ* ‘mare’ < **aspa-*, etc. In modern Ossetic, *p* (unless geminated) is limited to borrowed words—e.g., *patcax* ‘king’ (Persian *pādīšāh*), *pec* ‘oven’ (Russian *peč*).

Table 35-1. Consonantal Phonemes

	Digor Dialect			Iron (Literary) Dialect		
	–Voice	+Voice	Glottalic	–Voice	+Voice	Glottalic
Labial						
Stop	p	b	pʼ	p (п)	b (б)	pʼ (пъ)
Continuant	f	v		f (ф)	v (в)	
Nasal		m			m (м)	
Apical						
Stop	t	d	tʼ	t (т)	d (д)	tʼ (тъ)
Affricate 1	c	ʒ	cʼ	c (ц)	ʒ (дж)	cʼ (цъ)
Affricate 2				č (ч)	ž (дж)	čʼ (чъ)
Sibilant	s	z		s (с)	z (з)	
[Sibilant 2] ^a				š (dial.)	ž (dial.)	
Nasal		n			n (н)	
Liquid		r			r (р)	
Lateral						
Liquid		l			l (л)	
Velar						
Stop	k	g	kʼ	k (к)	g (г)	kʼ (къ)
[Labiovelar] ^a						
Stop	k ^w	g ^w	k ^{wʼ}	k ^w (кy)	g ^w (gy)	k ^{wʼ} (къy)
Uvular						
Stop	q			q (хъ)		
Continuant	x	ɣ		x (х)	ɣ (гъ)	
[Labiouvular] ^a						
Stop	q ^w			q ^w (хъy)		
Continuant	x ^w	ɣ ^w		x ^w (хy)	ɣ ^w (гъy)	

^aThe second sibilant series (š, ž) is limited to certain South Ossetian dialects of Iron (see below); the phonemic status of the labiovelars (k^w, g^w, k^{wʼ}) and the labiouvulars (q^w, x^w, ɣ^w) continues to be debated.

‘sheepskin’ (**wær*- ‘lamb’ (cf. Iron *wærikk* id.) + *carm* ‘skin’, with dissimilation of **r* > *l*). Neologistic compounds, however, often retain the voicelessness and aspiration of the stop, particularly in careful speech (Sokolova 1953: 37–39).

Table 35-2. Semivocalic Phonemes

	Digor	Iron
Palatal	j	j (ǰ)
Labiovelar	w	w (y)

Table 35-3. Vocalic Phonemes

	Digor			Iron		
	Front	Central	Back	Front	Central	Back
High	i [ī] ^a		u	i (ɨ) i (ɤ)		u (y)
Mid	e		o	e (e)		o (o)
Low		æ	a		æ (æ)	a (a)

^aThe phonemic status of Digor *ī* remains open to debate (see § 35.5.4).

A fully voiced pronunciation for the voiced occlusives *b*, *d*, *ʒ*, *g*, and (in Iron) *ǰ* is typical only medially after sonorants and in intervocalic position.³ Elsewhere these phonemes are normally either partially or completely devoiced, their distinguishing characteristic being the absence of aspiration. They are invariably voiceless when geminated (Iron *cittæn* ‘I was going’ [also spelled *ciddæn*, *cidtæn*] vs. *cidis* ‘he was going’), merging in this situation with the geminated voiceless phonemes (Sokolova 1953: 39–41).

Elements of the glottalic ejective series (*pʼ*, *tʼ*, *cʼ*, *kʼ*, and Iron *čʼ*) occur with less frequency in Ossetic than in neighboring Caucasian languages. The

3. Sokolova adds to the environments associated with complete voicing of the phonemes of this series “after voiced fricative consonants” (Sokolova 1953: 39), and cites Digor *tuydon* ‘military’. This appears to be true as long as the continuant is voiced. It is not uncommon, however, for clusters consisting of voiced continuant + voiced stop to be written (and presumably pronounced) without voicing—cf., e.g., *zaxton* alongside *zaydon* ‘I said’ (inf. *zæy-in*). This provides further evidence for Sokolova’s claim that voicing is not a distinctive feature of this series.

majority of the forms in which they occur are Caucasian loanwords—e.g., Iron *k'utær*, Digor *k'otær* ‘bush’ < Čečen *k'otær*, *žit'ri* ‘cucumber’ < Georgian *k'it'ri*, Iron *k'ux*, Digor *k'ox* ‘hand’ (origin unknown). Glottalic phonemes also occur in Russian loanwords, where they reflect Russian voiceless (un-aspirated) stops—*bulk'on* ‘colonel’ < *polkovnik*, *bap'iroz* ‘cigarette’ < *papiros* (Abaev 1949: 518–25). Sokolova (1953: 36–37) characterizes the distinctive component of the glottalic phonemes as a pause of .03–.04 second before the release of the vocal cords and the onset of a following vowel.

It is possible to draw a familiar analogy between the aspirates and ejectives: for both types of articulation the explosion does not signify the end of the consonant, and the vowel does not begin immediately after the explosion. But with the aspirates the explosion gives way to aspiration (the noise of the exiting airstream), but with the ejectives—to a pause. (Sokolova 1953: 37)

Abaev (1949: 512–17), on the basis of work by Axvlediani (1924), has identified a “fourth series” of occlusives *p*, *t*, *c*, *č*, *k* in Ossetic, distinguished from the voiceless phonemes described above by virtue of their lack of aspiration. He describes the distribution of these elements as restricted to (1) the position following a voiceless continuant (*s*, *f*, *x*), and (2) in gemination.

	(1)	(2)
<i>p</i>	—	<i>æppæt</i> ‘all’
<i>t</i>	<i>isti</i> ‘anything’	<i>cittæn</i> ‘I went’
<i>c</i>	<i>fæscæwin</i> ‘servant’	<i>liccinæn</i> ‘I shall run’
<i>č</i>	<i>isči</i> ‘anybody’	<i>wæriččitæ</i> ‘lambs’
<i>k</i>	<i>xæskæ</i> ‘bearing’	<i>wærikk</i> ‘lamb’

Although, as Abaev observed, “The consonants of the 4th series cannot be called phonemes in the full sense in Ossetic” (1949: 517), the role of these elements in Ossetic phonology remains subject to debate. Karažaeв (1984: 89–97) regards them as archiphonemes of the voiced and voiceless series, composed of the distinctive features (occlusion and point of articulation) common to both series.

Thus (*k*) and (*g*) (as well as the remaining stops) after (*S*), (*X*), and (*F*) undergo assimilation in the area of the constant [distinctive features], as the result of which there takes place a neutralization of their variable [distinctive features]. As a result, to these two phonemes (*k*) and (*g*) (as well as to the others) there corresponds in this position a third element *K*, which ... in linguistics is called an archiphoneme. (Karažaeв 1984: 96)

Sokolova (1953: 41–42), on the other hand, regards the “fourth series” as “one of the main variants of the Ossetic voiced consonants, which are characterized by their incomplete voicing.” Voicing, for Sokolova, cannot be

treated as the fundamental distinctive characteristic of the series *b*, *d*, *ʒ*, (Iron) *ʒ*, *g*, since these sounds so consistently occur without it. If the members of this series are considered to be basically “unaspirated” rather than “voiced,” Abaev’s fourth series becomes simply a set of allophonic counterparts to the “voiced” phonemes.

As has been indicated above, a morpheme boundary may cause the retention of the aspiration of a voiceless stop (*fīs-t^(h)æ* ‘sheep (pl.)’, with aspirated [^h]). Moreover, as Abaev observes (1949: 515), other instances of aspiration after *f*, *s*, *x* are also to be found—*kæxc^(h)* ‘large bowl’, *æxc^(h)æ* ‘money’, *Mæsk^(h)i* ‘Moscow’. In light of the existence of such forms, it might be better to assume that the neutralization of the voiceless and the voiced series after *f*, *s*, *x* is at best only partial, since only stops which are underlyingly voiceless, it appears, have the option of retaining the aspiration. Contrast the nominal plural suffix (with underlying voiceless *t*) and the past-stem formant (with underlying *d*).

	After a voiceless continuant	After a vowel
Plural <i>-t(æ)</i>	<i>fīs</i> -[<i>t^(h)</i>]æ ‘sheep (pl.)’	<i>xwɪ</i> -[<i>t^(h)</i>]æ ‘pigs’
Past-stem <i>-d</i>	<i>xas</i> -[<i>t^(h)</i>]-on ‘I carried’	<i>za</i> -[<i>d</i>]-is ‘it grew’

For instances such as these, therefore, it is reasonable to follow Sokolova in interpreting the elements of the “fourth series” as allophones of the “voiced” series (i.e., underlying /*fistæ*/ vs. /*xasdon*/), with voicing absent in the environment of the preceding voiceless consonant. It is not inconceivable that such an interpretation could be applied generally throughout the language (e.g., *færsk* ‘rib’ ← underlying /*færsg*/ vs. *æfc^(h)æg* ‘neck’ ← /*æfcæg*/). On the other hand, the invariable presence of the characteristics of the “fourth series” (i.e., the lack of aspiration and voicing) among the geminated obstruents, even in those cases in which the underlying obstruent is clearly of the voiceless series (e.g., *æmbæI-ttæ*, plural of *æmbal* ‘companion’), may well point to an archiphonemic analysis such as that proposed by Karažev.

35.4.1.1. Gemination

In addition to being found in a large number of lexical forms, gemination is systematically produced by morphophonological rule in several areas of the grammar.

(1) The first consonant of a verb stem is geminated when preceded by the preverbal prefix *ni-* (Digor *ni-*), which marks perfectivity and, with certain verbs, the notion “downward”: Iron *niwwaʒ-in*, Digor *niwwaʒ-un*

Table 35-4. Transitive Past Tense

	Singular		Plural	
	Iron	Digor	Iron	Digor
1	amardton	ramardton	amardtām	ramardtān
2	amardtaj	ramardtaj	amardtāt	ramardtājtæ
3	amardta	ramardta	amardtoj	ramardtoncæ

‘leave (inf.)’ (imperfective *waʒ-in*, *waʒ-un*), Iron *niccæw-in*, Digor *niccæw-un* ‘descend (inf.)’ (vs. *cæw-in*, *cæw-un* ‘go’). In Digor the preverb *fæ-* (marking perfectivity, among other features) exercises the same geminating effect on the first consonant of the verb-stem, while the corresponding Iron preverb has no such effect: Digor *fæccæw-un* ‘depart’ vs. Iron *fæcæw-in*. The gemination associated with these preverbs also influences the epenthetic segment (Iron *-c-*, Digor *-j-*) inserted between a preverb and a vowel-initial verb-stem: Iron *niccaxs-in*, Digor *nijjaxs-un* ‘catch (perf.)’ (imperfective *axs-in*, *axs-un*), Digor *fæjjarxaj-un* ‘try’ (vs. Iron *fæcarxaj-in*) (imperfective *arxaj-in*, *arxaj-un*). Note that both of these preverbs end in a weak vowel; the remaining preverbs, all of which end either in a strong vowel (*a-* (Iron), *ba-*, *æba-*, *ra-*) or in a consonant (*æ-*, Iron *(i)s-*/Digor *(i)s-*), do not cause gemination.⁴

(2) The *t* of the nominal plural marker *-tæ* is frequently geminated following a stem ending in a sonorant (*m*, *n*, *ɾ*, *l*) or a semivowel (*w*, *j*). The appearance of the gemination is closely connected with the weakening of the stem-vowel *a* (and *o* before nasals) to *æ* (*aryaw* ‘tale’, pl. *aryæw-ttæ*; *kom* ‘ravine’, pl. *kæm-ttæ*; *æmbal* ‘companion’, pl. *æmbæl-ttæ*).⁵ The gemination of the plural *-t-* does not take place when there is no weakening of the stem-vowel (*kaw* ‘fence’, pl. *kaw-tæ*), nor is it found when the stem-vowel is underlyingly weak (*zæj* ‘avalanche’, *zæj-tæ*).⁶

(3) The *d* of the past-stem formant is geminated before the transitive past-tense endings (cf. Table 35-4, the paradigm of *amar-in*, Digor *ramar-un* ‘kill’).

4. The gemination is found only in the case of productive preverbs. A small number of verbs contain an etymological **ni-* with no gemination (Iron *nig^uil-in* ‘bend down,’ *nwaz-in* ‘drink.’)

5. In a few words Digor shows the gemination where Iron does not—e.g., *bon* ‘day,’ pl. *bænttæ* (= Iron *bon*, *bon-tæ*, without weakening); *yæʒol* ‘blanket,’ pl. *yæʒælttæ* (= Iron *qæʒul*, *qæʒul-tæ*); *æfsinæ* ‘mistress,’ pl. *æfsinttæ* (= Iron *æfsin*, *æfsintæ*) (Abaev 1949: 384, Isaev 1966: 34–35).

6. In Iron a stem-final *-n* disappears by assimilation before the plural marker—*ron* ‘belt’, pl. *ræ-ttæ*, *don* ‘river’, pl. *dæ-ttæ*.

This gemination remains unexplained, as does much of the formation of the transitive past.

For intransitive verbs in the past tense, which are formed by suffixing the present tense forms of the verb ‘be’ to the past stem, gemination is found only in the first and second persons singular, where it evidently results from the juncture of the past-stem formant *-d* and an initial *d-* (of unclear origin) found in the corresponding forms of the copula (*dæn* ‘(I) am’ < Iran. **ahmi*, *dæ* ‘(you sg.) are’ < Iran. **ahi*). Cf. the past tense of *cær-in*, *cær-un* ‘live’ (past stem *card*), Table 35-5.⁷

Table 35-5. Intransitive Past Tense

	Singular		Plural	
	Iron	Digor	Iron	Digor
1	cardtæn		cardistæm	cardan
2	cardtæ		cardistut	cardajtæ
3	card(is)	cardæj	cardisti	cardæncæ

35.4.2. Sibilants and Affricates

In early Ossetic the distinction between the “hissing” and “hushing” sibilants of early Iranian (**s*, **z* and **š*, **ž*) was lost (*ars* ‘bear’ < Iranian **ṛša-*, cf. Avestan *arša-*, Skt. *ṛkṣa-*, Gk. *αρκτος*, Iron *mizd* ‘payment’ < **mižda-*, cf. Avestan *mižda-*, Skt. *mīḍha-*, Eng. *meed*). Similarly, Iranian **č* and **ǰ* have become *c* and *ǰ* (Iron *cæw-in* ‘go (inf.)’ < Iranian **čyav-*, cf. Avestan *šyav-*, Skt. *cyav-*, Iron *tinǰ-in*, Digor *itinǰ-un* ‘stretch’ < **vi-θanǰ-*, cf. Modern Persian *āhanjīdan*, Avestan *θanjaya-*).

New “hushing” sounds (*š*, *ž*, *č*, *ǰ*) have arisen in Digor as allophonic variants of *s*, *z*, *c*, *ǰ* in the environment of front vowels (Digor *čiyd* ‘cheese’, *æxsir* ‘milk’ vs. Iron *cixt*, *æxsir*). Unrelated to these Digor allophones are the Iron phonemes *č*, *ž*, *č’*, which have developed by palatalization from the velars *k*, *g*, *k’*. In the Dzaw dialect of South Ossetia (Tybylty 1929: 7, Abaev 1949: 496–97) a new sibilant series *š* and *ž* has appeared alongside Common

7. The discrepancies seen between Iron and Digor in the formation of the intransitive past paradigm reflect discrepancies in the forms of the verb ‘be’ in the two dialects:

1 sg. <i>dæn</i>	1 pl. Iron (<i>i</i>) <i>stæm</i> /Digor <i>an</i>
2 sg. <i>dæ</i>	2 pl. Iron (<i>i</i>) <i>stut</i> /Digor <i>ajtæ</i>
3 sg. Iron <i>u</i> , <i>i</i> , <i>is</i> /Digor <i>æj</i> , <i>jes</i>	3 pl. Iron (<i>i</i>) <i>sti</i> /Digor <i>æncæ</i>

Ossetic *s* and *z* through the simplification of the affricates *c* and *ʒ* (*š* ‘what’, *žur* ‘say (inf.)’ = Lit. Iron *ci*, *žurin*, Digor *či* (where *č* is the palatalized allophone of *c*), *žorun*).⁸

35.4.3. Labiovelars

35.4.3.1. Digor

The number of consonants ascribed to Ossetic by a given phonemic analysis depends upon the status assigned to the labiovelar phones. Early Iranian had a **xʷ* (< Indo-European **sw*), which still survives in the Digor dialect before back vowels (e.g., Digor *xʷæræ* ‘sister’ < Indo-European **swes(o)r-*), the labial component of **xʷ* having been lost before front vowels in Common Ossetic (Iron *xid*, Digor *xed* ‘sweat’ < earlier **xʷēd-* < early Iranian **xʷaid-*, cf. Avestan *xwaēda-*, Skt. *sveda-*). New labialized counterparts to the other velars and uvulars have been introduced into the language through loanwords—e.g., *læqʷæn* ‘boy’ (< Oss. *læg* ‘man’ + Northwest Caucasian *qwa* ‘son’), *kʷære* ‘week’ (< Georgian *kʷira* ‘week, Sunday’, ult. < Greek *κυριακή*)—leading to the appearance of what is arguably a fairly complete set of rounded back consonants. These consonants lose their rounded quality before rounded vowels (cf. *xonx* ‘mountain’ vs. pl. *xʷænx-tæ*).

For Digor, the issue of the phonemic status of the labiovelars takes the form of a debate over whether the sounds in question should be thought of as phonemic units or as sequences of velar/uvular + semivowel *w*.⁹ Sokolova observes that, although phonetic work remains to be done in this area, the “*a* of Digor *xʷar* ‘bread’ and the *wa* of *dwar* ‘door’ appear to give different acoustical impressions.

1) In sequences of the type of *dwar* an actual diphthong, with a strong first element, is heard (“*duar*”); in sequences with *x*, the first element is weaker and shorter (“*xʷar*”). Diphthongization of this sort is normal with a preceding labialized consonant; consequently, the phonological sequence *xʷa* must give exactly the sound “*xʷa*.” 2) With a sequence of the type of *dwa* the vowel does not alter its quality, maintaining its constant sound (“*duar*”). But in the sequence *xʷa*, *xʷə* [= *xʷa*, *xʷə*—D.T.] the vowel changes, receiving a slight roundedness. A more conspicuous change is shown by *a*, which comes close to an “*o*”-like sound, particularly in faster speech (“*xʷor*”). (Sokolova 1953: 48)

8. This is a quite recent development. Material on this dialect from the first half of the nineteenth century renders the Common Ossetic affricates as *č* and *ž*, evidently the midpoint in the evolution from *c*, *ʒ* to *š*, *ž* (Abaev 1949: 496–97).

9. Isaev (1966: 18), citing Sokolova (1953), does not list the labialized velars and uvulars among the phonemes of Digor, whereas Thordarson (1989: 462) does.

Further research will be required to resolve the issue of the status of the labiovelars and labiouvulars in Digor.

35.4.3.2. *Iron*

The Iron dialect is characterized by the merging of the early Ossetic high short vowels **i* and **u* into *i*. When preceded by a velar or uvular, however, the rounded quality of an earlier **u* was transferred to the consonant.¹⁰ As a result, each of the velar consonants (*k*, *g*, *k'*) and the uvular consonants (*q*, *x*, *y*) has acquired a labialized counterpart, found only before the vowel *i*.¹¹

Digor	Iron	
<i>kud</i>	<i>k^wid</i>	'how'
<i>igurun</i>	<i>g^wirⁱⁿ</i>	'be born'
<i>mæk'ustæg</i>	<i>mæk^wistæg</i>	'shoulderbone'
<i>quran</i>	<i>q^wiran</i>	'Qur'ān'
<i>mægur</i>	<i>mæg^wir</i>	'poor'
<i>xumæ</i>	<i>x^wim</i>	'field'
<i>wæryun</i>	<i>wæry^win</i>	'lamb's wool felt'

These elements have given rise to a considerable amount of debate in discussions of the synchronic phonology of Ossetic, and various investigators have chosen differing approaches to these elements. Miller (1962: 38) in effect took the rounded feature to be still inherent in the vowel, claiming that the earlier **u* has survived in post-velar position as *ũ*. Sokolova, recognizing that the quality of the vowel of *k^wi* differs from that of *i* elsewhere, states that "phonetically there is heard a diphthong with an alteration of the quality of *i* in the direction of 'u', so that one often hears not 'k^uid' but 'k^uød', which led Miller to find in these cases a short *u*" (Sokolova 1953: 49).

Another approach is to view *k^wi* as fundamentally *k* + *i*, the labial component of the sequence being ascribed to a secondary phonological process affecting sequences of velars (or uvulars) and the high central vowel. This approach may be found outlined in Karažev (1984: 85–89), where it is argued that the labial character of these elements is to be ascribed to a form of epenthesis which seeks to accommodate the difference between the velar

10. Since early Ossetic velars (**k*, **g*, **k'*) became affricated (*č*, *ž*, *č'*) before **i* in Iron (e.g., *čizg* 'daughter' vs. Digor *kizgæ*), in principle one might expect there to be no sequences of the type *k^wi*, *g^wi*, *k^wi* (i.e., without either palatalization (before **i*) or labialization (before **u*)) surviving in native words of the language.

11. This labialized series either has not developed or has been lost in the dialect of the Alagir ravine, where one finds, e.g., *x^wi* rather than *x^wi^w* 'pig,' *k^wivd* rather than *k^wi^wvd* 'prayer' (Karažev 1984: 89 n. 5).

(or postvelar) point of articulation of the consonants and the high central articulation of the vowel *i*. The intrusive *w*, being simultaneously high like *i* and articulated “like a sound of the back series (by which, incidentally, is explained the fact that the labialized sounds are not palatalized before ‘*i*’)” (Karažaev 1984: 88), serves as a bridge between the two. The labialized velars and uvulars are thus not phonemically distinct from the plain velars and uvulars.¹²

Karažaev’s argument is vitiated by the fact that Iron does in fact allow sequences of velar/uvular + *i* in which the *w* fails to appear. Leaving aside children’s words such as *gigi* ‘girl’, the expected *w* fails to surface before the adessive case-marker *-il*—e.g., *læg-il* ‘on the man’.¹³ In this suffix the vowel *i* is historically a secondary development (cf. the Digor adessive *læg-bæl*, the suffix ultimately coming from Iranian **upari*), a fact which suggests that, while Karažaev’s *w*-insertion may have taken place at one point in time, it can no longer be regarded as an automatic phonological process (note also the morphophonemic nature of the rule producing labiovelar *-gʷ-* before the plural marker *-tæ*, discussed in § 35.4.4).

It is interesting to observe that the development of Iron has apparently involved a wholesale elimination of the labiovelars/labiouvulars outside of the environment of a following *i*. Digor, as we have seen above, employs a fairly extensive set of back rounded elements (whatever their phonemic status may be). The Iron counterparts to Digor words containing these sounds indicate that the labiovelars/labiouvulars have been systematically eradicated, either through (a) complete loss of the labial quality or (b) transfer of the labial quality to an adjacent vowel.

	Digor	Iron	
(a)	<i>xʷærun</i>	<i>xærin</i>	‘eat’ (inf.)
	<i>xʷæcun</i>	<i>xæcin</i>	‘hold’ (inf.)
	<i>kʷæcæl</i>	<i>kʰæcæl</i>	‘sliver’
(b)	<i>xʷærxæ</i>	<i>xo</i>	‘sister’
	<i>xʷarz</i>	<i>xorz</i>	‘good’
	<i>kʷassæ</i>	<i>kosa</i>	‘beardless’

It therefore seems that there has been a series of changes eliminating and subsequently reinstating the rounded back consonants in the history of Ossetic, and particularly of Iron. At an early stage of Ossetic the **xʷ*, the sole

12. This is the view expressed in Abaev’s 1952 sketch (p. 447).

13. There is a measure of dialectal variation in the issue of the adessive ending, however, since forms such as *lægʷil* are not unknown. The phonological status of the labiovelars and labiouvulars may well vary from dialect to dialect.

native Iranian representative of this class, was lost before front vowels. An influx of loanwords introduced new words containing labiovelars and labio-uvulars, but in early Iron all of these back rounded consonants seem to have been eliminated. Subsequently Iron created a new series of labiovelars and labio-uvulars through the merger of earlier **u* with **i* as *i* and the transfer of the labial quality of the **u* to a preceding back consonant.

35.4.4. Palatalization (Iron)

It has been seen above that the common Ossetic affricates *c* and *ɟ* have developed from early Iranian **č* and **ǰ* by a shift in the point of articulation which is also seen in early Iranian **š* > Ossetic *s*. However, a new set of affricates *č*, *ǰ*, *č'* has developed in the Iron dialect as the result of a palatalization of the velars *k*, *g*, *k'* under the influence of a following front vowel. Once the merger of the vowels **i* and **u* into Iron *i* took place, the distinction between the velars and the new affricates became phonologized. In the following examples the older shape of the velars is retained in the Digor forms, while the Iron forms show the affricates arising through palatalization.

Digor	Iron	
<i>kinzæ</i>	<i>činz</i>	'bride'
<i>kiunugæ</i>	<i>činig</i>	'book'
<i>ragi</i>	<i>raǰi</i>	'early, long ago'
<i>gepp</i>	<i>ǰipp</i>	'mold'
<i>ærk'e</i>	<i>ærč'i</i>	'rawhide boot'
<i>k'ildun</i>	<i>č'ildim</i>	'back, rear'

The palatalization process is still a live phenomenon in modern Iron, as may be seen, for example, in alternations within the nominal paradigm—*læɡ/læǰ-i* 'man (nom. sg./gen. sg.)', *kark/karč-i* 'chicken (nom. sg./gen. sg.)'. It is doubtful, however, that this palatalization is still of a purely phonological nature. The genitive ending *-i*, for example, may be contrasted with the comitative ending, which, despite the presence of a front vowel, does not normally cause palatalization of a preceding velar¹⁴—*læɡ-imæ* 'with the man'. Palatalization is likewise absent from a number of lexical items containing a sequence of velar + front vowel: *k'en* 'checkers (children's word)', *gino* 'kitty (children's word)', *gigienæ* 'hygiene (< Russian)' (Karažaev 1984: 98–105). In light of the existence of such unpalatalized velars, it is probably

14. The comitative case in *-imæ* is absent from Digor, where the corresponding construction is formed with the postposition *xæccæ* (*læɡ-i xæccæ*). This suggests that the comitative case arose as an Iron innovation after palatalization had ceased to be a purely phonological phenomenon.

best to regard palatalization in the modern language as a grammatical process associated with the morphology of case and number marking.¹⁵

An interesting morphophonological phenomenon is to be found in connection with the plural formant *-tæ* in Iron. This suffix requires the introduction of a linking vowel *-i-* for the majority of noun stems ending with a consonant cluster—*mist* ‘mouse’, pl. *mist-i-tæ*, *xæfs* ‘frog’, pl. *xæfs-i-tæ*, *binz* ‘fly’, pl. *binz-i-tæ*. As may be seen from examples such as *čizg* ‘girl’, pl. *čizž-i-tæ*, and *dærk* ‘kid’, pl. *dærč'-i-tæ*, the inserted vowel has the effect of palatalizing a preceding velar. The vowel insertion in question also occurs (aside from a few exceptions) with stems ending in the suffixes *-æg* and *-ig*, in which case there is syncope of the vowel preceding the *-g-*, unless the resulting consonant cluster would be disallowed by the phonotactics of the language—*zuræg* ‘speaking’ (participle of *zur-in*), pl. *zurž-i-tæ* vs. *xæjræg* ‘demon’, pl. *xæjræž-i-tæ*. The syncope of *-ig*, however, has important repercussions for the quality of the stem-final velar.

+syncope	-syncope
<i>-æg</i> <i>zuræg</i> ‘speaking’, <i>zurž-i-tæ</i>	<i>xæjræg</i> ‘demon’, pl. <i>xæjræž-i-tæ</i>
<i>wazæg</i> ‘rooster’, <i>wazž-i-tæ</i>	<i>sintæg</i> ‘bed’, <i>sintæž-i-tæ</i>
<i>baræg</i> ‘rider’, <i>barž-i-tæ</i>	<i>qoppæg</i> ‘lily’, <i>qoppæž-i-tæ</i>
<i>-ig</i> <i>mæsig</i> ‘tower’, <i>mæsg^w-i-tæ</i>	<i>mæłzig</i> ‘ant’, <i>mæłziž-i-tæ</i>
<i>činig</i> ‘book’, <i>čing^w-i-tæ</i>	<i>qæzdiğ</i> ‘rich’, <i>qæzdiž-i-tæ</i>
<i>wæjig</i> ‘giant’, <i>wæjg^w-i-tæ</i>	<i>færniğ</i> ‘happy’, <i>færniž-i-tæ</i>

The syncope of *-ig* has as a by-product the replacement of the velar with a labiovelar *-g^w-* (rather than palatalized **-ž-* as in *barž-i-tæ*). Note that the labialization of the velar is contingent upon the loss of the vowel (contrast *mæłziž-i-tæ*, in which syncope is blocked by the consonant cluster and the velar is palatalized). It is clear from the Digor counterparts to the Iron forms ending in *-ig* (Digor *mæsug*, *mulzug*, etc.) that the original vowel of this suffix was **u*, which at an earlier point in history was evidently metathesized with the following velar (**mæsug-tæ* > **mæsgu-tæ* > *mæsg^w-i-tæ*).

35.4.5. Uvular *q*

In Digor, *q* is essentially restricted to loanwords from Turkic (*qumac* ‘fabric’, *basluq* ‘bashlyk’, *qæma* ‘dagger’) (Isaev 1966: 15). Iron, in addition to showing such forms (cf. *q^wimac*, *basiliq*, *qama*), has expanded the range of its *q* by the shift of earlier **ɣ* in initial position to the voiceless stop *q*—cf. Iron *qus*

15. Palatalization is also found (optionally) in the declension of Russian loanwords (e.g., *uroč-itæ*, plural of *urok* ‘lesson’).

‘ear’ vs. Digor *yos* (< Iranian **gauša-*), Iron *qəbatir* ‘brave’ (by metathesis from **bəyatur-*; cf. Iron *bəyatir*, Digor *bəyatær* ‘hero’).

35.4.6. *Nasals and Liquids*

An early Ossetic final *-*m* sometimes takes the form -*n* in Digor—Digor *non* vs. Iron *nom* ‘name’, Digor *cəwæn* vs. Iron *cəwæm* ‘we go’. In Iron, **n* has in many cases been completely assimilated to a following consonant—Iron *fissin* vs. Digor *finsun* ‘write (inf.)’, Iron *ssæʒ* vs. Digor *insæj* ‘twenty’ (cf. Skt. *vimśati-*). Several Digor forms show the effects of metathesis in clusters consisting of a sibilant + a nasal: Digor *fans* vs. Iron *fæsm* ‘wool from the autumn shearing’, Digor *anz* vs. Iron *az* (< **azm-*) ‘year’. Final -*m* is lost in several Digor forms after *r* (Digor *yar* vs. Iron *qarm* ‘warm’), and final -*n* is lost in Iron after *oj* (Iron *moj* vs. Digor *mojnæ* ‘husband’, Iron *k^wiroj* vs. Digor *kurojnæ* ‘mill’).¹⁶ Iron often retains *m* unassimilated to a following consonant in words whose Digor counterparts show *n* or *ŋ*: Iron *simd* ‘group dance’, *æmxæræfirt* ‘kinsman with the same maternal uncle’ vs. Digor *sind*, *æŋx^wærifurt*.

35.4.7. *Semivowels*

The semivowels *j* and *w* occur in both dialects (Iron/Digor *cəjaw* equative case of *ci/ci* ‘what’, Iron *cəwin*, Digor *cəw(u)n* ‘go (inf.)’). In Iron the sequences -*awæ-* and -*əwa-* are simplified to -*ua-* (Iron *cuan* vs. Digor *cawæn*

16. One of the clearest differences between the shape of words in Digor and Iron is the widespread presence in the former dialect of a final vowel -*æ* which is absent from Iron—cf. Iron *fid* vs. Digor *fidæ* ‘father,’ Iron *mist* vs. Digor *mistæ* ‘mouse,’ Iron *xicawad* vs. Digor *xecawadæ* ‘government,’ Iron *cærinç* vs. *cæruncæ* ‘live (3 pl. indic.),’ etc.—although in many forms the final -*æ* is found in both dialects—cf., inter alia, Iron/Digor *bæx-tæ* ‘horses’. The original presence of *-*æ* in pre-Iron words for which Digor shows -*æ* is indicated by the distinction between the vowels *a* and *æ* in Ossetic stems reflecting the original shape *-*aCC-*: as part of a monosyllable, pre-Ossetic *-*aCC* normally went to -*aCC*, whereas elsewhere it yielded -*æCC*—cf. Iron *qarm* ‘warm’, *bast* ‘bound’ (= Digor *yar(m)*, *bast*) vs. Iron *wærm* ‘pit,’ *bæstæ* ‘land’ (= Digor *wærmæ*, *bæstæ*).

	*-aCC	*-aCCV-	*-aCCV-
Iron	<i>qarm</i>	<i>wærm</i>	<i>bæstæ</i>
Digor	<i>yar(m)</i>	<i>wærmæ</i>	<i>bæstæ</i>

The presence of the vowel *æ* rather than *a* in Iron forms such as *wærm* thus indicates an early Ossetic bisyllabic stem-shape, agreeing with the Digor stem-shape in -*æ*. In the case of Iron -*oj* vs. Digor -*ojnæ*, we find the familiar loss of -*æ* in Iron followed by the loss of the (now-final) *-*n* after *-*oj-*.

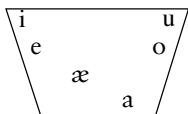
‘hunt’, Iron *quag* vs. Digor *yæwagæ* ‘shortage’), while *-æwæ-* is typically simplified to *-o-* (Iron *nog* vs. Digor *næwæg* ‘new’).¹⁷ In initial position Digor *w* sometimes appears as *j* before a front vowel—Digor *wes/jes* vs. Iron *wis* ‘twig’, Digor *winun/jinun* vs. Iron *winin* ‘see (inf.)’.

New instances of *j* have arisen (evidently independently) in each dialect from earlier affricates in word-final position—Digor *kuj* vs. Iron *kʰɯʒ* ‘dog’, Digor *fij/fiŋzæ* vs. Iron *fiŋz* (see below), Iron *farstoj* vs. Digor *farstoncæ* ‘they asked’ (Abaev 1968).

35.5. The Vowel System

35.5.1. Digor

The vocalic system of Digor Ossetic, which preserves several archaic features no longer found in Iron, consists of the following:



Of these, *i*, *u*, and *æ* are the “weak” vowels, characterized by shorter length and a greater degree of allophonic variation. The “strong” vowels *e*, *o*, and *a*, on the other hand, are primarily descended from the earlier diphthongs **ai* and **au* and the long vowel **ā*, respectively; many instances of *o* (in both Digor and Iron—see § 35.5.3) may, in addition, be traced back to earlier **ā* before nasals (e.g., Digor *non*, Iron *nom* ‘name’ < **nāman-*), although this shift appears to be a rather late change. The distinction between weak and strong vowels is crucial for the placement of stress.

In terms of its realization, Digor *i* is a short front vowel, the height of which ranges from quite high (in the diphthong *ij*—see below) to a lax *ɪ* (e.g., in the vicinity of sibilants) and even to a close *e* (e.g., in final position or before uvulars). The range of its back, rounded counterpart *u* is narrower. The *æ* is a central, rather low vowel not far from a shwa, with a tendency towards frontness (*æ*) in the diphthong *æj* and towards backness in the environment of uvulars or *w*.

The strong vowels are all typically longer than the weak vowels. The *e* and the *o* are pronounced quite high, the *o* even sounding like *u* before nasals; in word-initial position *e* and *o* are pronounced with a slight semi-

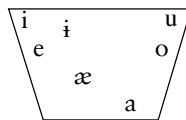
17. The Dzaw dialect of South Ossetia, although a subdialect of Iron, preserves *-æwæ-*: *næwæg* = literary *nog* ‘new,’ *ræwæd* = literary *rod* ‘calf’ (Tyblyty 1929: 6–7).

vocalic glide preceding (*ʼex* ‘ice’, *ʷosæ* ‘woman’). The *a* is low and produced quite far back.

35.5.2. Iron

The Iron vocalic system has undergone several modifications. The vowels *i* and *u* preserved in Digor have merged in Iron and are reflected in *ɨ* (Iron *fɨd* ‘father’, *miɨd* ‘honey’ vs. Digor *fiɖæ*, *mud*, cf. Avestan *pitā*, *madu*), a high central unrounded vowel (with a wide range of allophonic variation, including a fronted allophone in the vicinity of palatals and a rounded allophone in labial environments). A new pair of high vowels *i* and *u* has been created in Iron by the raising of earlier *e* and *o* (Iron *ævzist* ‘silver’, *qus* ‘ear’ vs. Digor *ævzestæ*, *ɣos*). The phoneme *o* resulting from **ā* before nasals has not been raised (*nom* ‘name’ = Digor *non*, Avestan *nāman-*), and new instances of *o* appear in Iron through the shift of *a* to *o* after labiovelars and labiouvulars (*xorz* ‘good’ vs. Digor *xʷarz*, *kʷord* ‘group’ vs. Digor *kʷar*) and of *-æwæ-* (Iron *nog* vs. Digor *næwæg* ‘new’); *e* has only a marginal existence in native Iron words, its occurrence being limited to a restricted number of cases resulting from the simplification of sequences of *j* and *æ* (e.g., *fæste* ‘behind’ alongside *fæstijæ*) and from vowel sandhi at morpheme boundaries (*me ʼfsimær* ‘my brother’ < *mæ* + *æfsimær*; *festin* ‘stand (perf.), become’ < aspectual prefix *fæ-* + (*i*)*stin* ‘stand (imperf.)’). The functional load of the mid-vowels has been dramatically bolstered by the introduction of loanwords (e.g., *sovet* ‘council’, *professi* ‘profession’).

The resulting set of vowels in Iron is the following:



As in Digor, the vowels are divided into weak (Iron *ɨ*, *æ*) and strong (Iron *a*, *i*, *u*, *e*, *o*) classes.

35.5.3. Summary of Vowel Correspondences

For discussion of the phonetic details of the realization of the vocalic phonemes in Digor and Iron see Sokolova (1953: 9–34). The basic vocalic correspondences between the two dialects are depicted in the following table.

Early Iranian	Digor	Iron
* <i>i</i> , * <i>ī</i> >	<i>i</i>	<i>ɨ</i>
* <i>u</i> , * <i>ū</i> >	<i>u</i>	<i>ɨ</i>
* <i>ay</i> >	<i>e</i>	<i>i</i>

* <i>aw</i> >	<i>o</i>	<i>u</i>
* <i>ā</i> >	<i>a</i>	<i>a</i>
* <i>āN</i> >	<i>oN</i>	<i>oN</i>
* <i>a</i> >	<i>æ</i>	<i>æ</i>
* <i>aCC</i> # ¹⁸	<i>aCC</i>	<i>aCC</i>

35.5.4. The Digor Vowel “*ī*”

Alongside the correspondences of Digor *i* with Iron *i* and Digor *e* with Iron *i*, Isaev (1954) has proposed recognizing a vowel correspondence represented by Digor *ī* and Iron *i*. Since there is already an *i* (< early Ossetic **e*) in Iron, the existence of this correspondence does not add to the phonemic inventory of this dialect. If it can be shown that there is a phonemic vowel *ī* distinct from *i* in Digor, however, it will be necessary to emend the vowel system accordingly. Isaev finds that the *ī/i* correspondence occurs either (a) in native Ossetic words before *n* (or, in some cases, in word-final position following the loss of an earlier *n*), or (b) in loanwords.

	Digor	Iron	
(a)	<i>cīnæ</i>	<i>cin</i>	‘joy’
	<i>æfsīnæ</i>	<i>æfsin</i>	‘mother-in-law, mistress’
	<i>suyzæīnæ</i>	<i>siyzærin</i>	‘gold’
	<i>sozīnæ</i>	<i>suzin</i>	‘needle’
	<i>fī</i> ¹⁹	<i>fīnʒ</i>	‘nose’
	<i>ærcī</i>	<i>ærcinʒ</i>	‘needle for sewing shoes’
(b)		<i>dīn</i>	‘religion’ (< Arabic)
		<i>ʒīn</i>	‘jinn’ (< Arabic)
		<i>klīmæt</i>	‘climate’ (< Russian)
		<i>gīmn</i>	‘hymn’ (< Russian)

Isaev adduces minimal pairs such as *dīn* ‘religion’ : *din* ‘you (dat. sg. enclitic)’, *æxsīnæ* ‘princess’ : *æxsinæ* ‘I shoot (optative)’ which leave little doubt that *ī* is distinct from *i* in Digor. It remains open to question, however, whether *ī* is best analyzed as a vocalic unit, or, as many researchers have chosen to view it, as a diphthong *ij* (i.e., short *i* + semivowel *j*). It is worth noting that taking *ī* to be *ij* relieves us of the necessity of explaining why *i* alone of all the vowels of Digor fails to appear in a *j*-diphthong. Certain of the instances of *ī* which Isaev presents clearly derive etymologically from diphthongization—it is clear from the Iron form *fīnʒ*, for example, that

18. See n. 16 above.

19. Beside the Digor by-form *fīnʒæ*; cf. also the frozen locative-case form *fīnʒi* ‘in front’.

Digor *fī* ‘nose’ must have developed as a result of a phenomenon parallel to that which gave rise to Digor *insæj* ‘twenty’, *kuj* ‘dog’ (vs. Iron *ssæʒ*, *kʷiʒ*)—i.e., **fīnʒ-* > **finj-* > **fijn-* > *fij*. Indeed, in a certain number of instances Digor *ī* appears to result synchronically from *i* via morphological processes which give rise to diphthongs elsewhere—cf. *adgīnag* ‘sweetness’ from *adgin* ‘sweet’ (Iron *ačīnag*, *ačīn*) alongside *xʷærujnag* ‘food’ from *xʷærun* ‘eat (inf.)’ (Iron *xærinag*, *xærin*), *bazajrag* ‘bazaar (adj.)’ from *bazar*.

35.5.5. Vowel Alternations

Alternations in the quality of stem vowels occur in a number of situations. For many words the addition of certain derivational suffixes produces a weakening of *a* to *æ* (*fars* ‘side’ vs. *færssag* ‘accessory’, *avd* ‘seven’ vs. Iron *ævdæm*, Digor *ævdæjmag* ‘seventh’). This weakening is also frequent in compound words (*ast* ‘eight’ vs. *æstdæs* ‘eighteen’), and is an important feature of the morphophonemics of plural marking in the nominal system—*kark* ‘chicken’, pl. *kærč-i-tæ*, *xox* (< **xonx* < **xanx*) ‘mountain’, pl. *xæx-tæ*.²⁰ In a number of verbal pairs, an opposition between strong vowel and weak vowel is employed to distinguish between transitive and intransitive stems—*saf-in* ‘lose’ vs. *sæf-in* ‘disappear’, *tag-in* ‘pour (trans.)’ vs. *tæʒ-in* ‘pour (intr.)’, *qus-in* ‘hear’ vs. *qʷis-in* ‘be heard’.

Vocalic alternation is important in the formation of the present and past verbal stems. Verbal stems containing a strong vowel *i* or *u* (Digor *e* or *o*) or a diphthong *-æw-* or *-aw-* in the present stem systematically replace that vowel with a weak vowel *i* (Digor *i* or *u*) in the past stem; present stems containing *æ*, on the other hand, typically show the strong vowel *a* in the past.²¹

In historical terms, the alternations *i/i* and *u/i* (= Digor *e/i* and *o/u*) reflect the distinction between an early Iranian full grade (e.g., Iron *cæw-*, *qus-* < **čyau-*, **vi-gauš-*) and a zero grade (*cid*, *qʷist* < **čyu-ta-*, **vi-guš-ta-*). This alternation has even been expanded in Ossetic to encompass the secondary *o* < **ā* before a nasal—cf., e.g., the present-stem *zon-* ‘know’ (< Iran. **zān-*), which has a past stem *zind* (Digor *zund*) as if from **zawn-*, rather than the expected **zond-* < Iran. **zan-ta-*.

20. This alternation reflects the development of early Ossetic short **a* into *a* in monosyllables ending in a consonant cluster—**aCC* > *-aCC*, as in *fars*—but into *æ* elsewhere—as in *færssag* (see n. 16).

21. In general the present stem of an Ossetic verb continues one of the various present-stem formations of Indo-Iranian (e.g., *mæł-in* ‘die’ < **mṛ-ya-*, *tæfs-in* ‘grow warm’ < **taf-sa-*, etc.). The past stem is based upon an earlier participle (*mard-* < **mṛ-ta-* ‘having died’). For discussion of the morphology of the Ossetic verbal system see Job (1982).

Table 35-6. Vowel Alternations in Verb Stems
(Iron above, Digor below)

	Present (3 sg. pres. indic.)	Past (3 sg. past indic.)
‘turn (intr.)’	zil-i	zil-d(-is)
	zel-uj	zil-d-æj
‘hear’	qus-i	q ^w is-t-a
	iyos-uj	iyus-t-a
‘go’	cæw-i	ci-d(-is)
	cæw-uj	cu-d-æj
‘sow’	taw-in	ti-tt-a
	taw-uj	tu-tt-a
‘live’	cær-in	car-d(-is)
	cær-uj	car-d-æj

The alternation *æ/a*, on the other hand, evidently results from the fact that early Ossetic short **a* took the shape *a* in monosyllabic words ending in **-CC* (see n. 16), rather than *æ* which is the reflex of short **a* elsewhere. The addition of the stem-suffix *-d* (< **-ta-*) to the verb-stem therefore led to the appearance of a vocalic distinction between the present and past stems—present-stem *cær-* ‘live’ vs. past-stem *card-*. In addition, other alternation-patterns have arisen dialectally, particularly in bisyllabic stems—Iron *xal-/xæld-* ‘destroy’ vs. Digor *ixal-/ixald-*, Iron *ærgævd-/ærgævst-* ‘slaughter’ vs. Digor *ævgærd-/ævgarst-* (with metathesis).²²

35.6. Phonotactics

Unlike many of the neighboring Caucasian languages, Ossetic has no great tolerance for complex consonant clusters. In loanwords from these languages, inserted vowels have broken up the clusters—e.g., Iron *cæʒinʒ*,

22. The vowel alternation rules as outlined here, it may be noted, have the effect of erasing the distinction between transitive and intransitive stems in the past tense in many verbs—e.g., *amard-*, past stem of both *amar-in* ‘I kill’ and *amæł-in* ‘I die’, *q^wist-*, past stem of both *qus-in* ‘I hear’ and *q^wis-in* ‘I am heard’. The task of distinguishing transitivity and intransitivity in the past is mainly borne by the verbal endings rather than the shape of the stem—*amartton* ‘I killed’, *amarttæn* ‘I died’, *q^wiston* ‘I heard’, *q^wistæn* ‘I was heard’.

Digor *cæginzæ* ‘post’ < Mingrelian **čxinǰi*,²³ Digor *cæqal* ‘wave’ < Georgian *c’qal-i* ‘water’. A prothetic vowel *æ* or *i* is introduced before consonant clusters in initial position (Iron *ærvad*, Digor *ærvadæ* ‘kinsman’ < Iranian **brātā* ‘brother’ (nom. sg.)). Certain morphophonological processes, such as the syncopation of the suffixes *-æg* and *-ig* before the plural ending (see § 35.4.4), may be blocked in order to avoid unallowably ponderous consonant sequences.

35.7. The Accent System

35.7.1. Accent Placement and Accent Complexes

Accent placement is based upon the distinction between the strong vowels (*a*, *e*, *o*, and Iron *i* and *u*) and the weak vowels (*æ*, Iron *ɨ*, and Digor *i* and *u*). For the Iron dialect, the basic rule is that the accent will fall upon one of the first two syllables—the first syllable if it contains a strong vowel²⁴ and the second syllable if the first syllable contains a weak vowel (*ústitæ* ‘women’ vs. *čizžitæ* ‘girls’). Personal names are uniformly accented on the second syllable regardless of their vocalism (e.g., *Sozírigo*, *Satána*, *Batíraz*, names from the folk epic of the Narts).

In the case of a few words, such as *sístin* ‘stand’ (inf.) and *mísín* ‘compose, devise’ (inf.), the accent falls upon the first syllable despite the weakness of its vowel. This unexpected situation is due to the loss of an earlier initial vowel (cf. Digor *isistun*, *imisun*), on account of which the accent did in fact fall on the second syllable at an earlier stage of the language (**imísun* > *mísín*). This phenomenon has important consequences for the marking of determinedness in noun phrases (see § 35.7.2).

The Digor accent system is considerably less well understood than the Iron system. It employs the same distinction between vocalic strength and weakness, but the placement of the accent is determined by different rules. When strong vowels are present, the accent normally falls upon the rightmost strong vowel (*raxastón* ‘I brought’, *fælváræ* ‘the year before last’); in the absence of a strong vowel, the accent falls upon the rightmost syllable (*γuru-múq* ‘rough’, *næ tikís* ‘our cat’), although endings with the weak vowel *-æ* may only be accented in bisyllables (*fidæ* ‘father’, but *ær-min-cæýdæ* ‘play for me’) (Isaev 1966: 26–27).

23. Not attested, but reconstructed on the basis of Georgian *čxiri* ‘stick,’ Abkhaz *čxənǰ* ‘stick for suspending a cauldron’ (Abaev 1949: 526).

24. In a few cases, such as *saqát* ‘shortcoming’ (< Arabic *saqat*), in which both the first and the second syllables contain strong vowels, the accent falls on the second syllable.

The accent-placement rules discussed above do not operate upon words but upon accentual complexes which are typically larger than a single word, and are often of considerable length.²⁵ A string of modifiers and their head, for example, form a single accentual unit, as do a substantive and its post-position or an object and a following verb. Clitics (both proclitics and enclitics, the latter of which are drawn into “Wackernagel’s position” following the first element of the clause) are absorbed into the accentual complexes formed by neighboring elements. These accentual complexes are the objects of the accent-placement rules; it may be noted that the element within the accentual complex which is assigned the accent by the rules is not infrequently a clitic. Consider, for example, the sentence *nir tá sin æj kúrinc sæ čízžī* ‘now again they ask them for her, their daughter’, which consists of seven words (three of which are enclitics and one a proclitic) arranged into three accentual complexes. Within each of these complexes, the accent is placed in accordance with the vowel-strength rules described above.²⁶

	<i>nir</i>	<i>tá</i>	<i>sin</i>	<i>æj</i>
	now	again	them (dat.)	3 sg. pron.
Accentual complexes:	X	encl.	encl.	encl.
Accent placement:	wk.	syll. 2	...	...
	<i>kúrinc</i>	<i>sæ</i>	<i>čízžī</i>	
	they ask	their	daughter (def. dir. obj.)	
	X	procl.	X	
	strong V	wk.	syll. 2	
	in syll. 1			

Abaev (1939: 97) cites an example of a pair of sentences distinguished solely through the composition of the accentual complexes and thus the placement of the accent: *bæx næ qæwí máx* ‘we need a horse’ vs. *bæx næ qæwí máx* ‘we do not need a horse’. In the first, *næ* is an enclitic pronoun coreferential with the independent pronoun *máx* ‘us’, whereas the *næ* of the second sentence is a proclitic negative.

	<i>bæx</i>	<i>næ</i>	<i>qæwí</i>	<i>máx</i>
	horse	us	is needed	us
Accentual complexes:	X	encl.	X	X
Accent placement:	wk.	syll. 2	wk. V in syll. 1	1 syll.

25. Long accentual complexes often develop a secondary accent in addition to the primary accent.

26. Abaev 1939: 103. Note that *sæ čízžī* constitutes its own accentual complex on account of its postverbal position.

<i>bæx</i>	<i>næ</i>	<i>qæwi</i>	<i>máx</i>
horse	neg.	is needed	us
X	procl.	X	X
1 syll.	wk.	syll. 2	1 syll.

For a comprehensive discussion of the accentual system of Ossetic see Abaev (1939, 1949: 529–60).

35.7.2. Accent and Definiteness (Iron)

We have seen above that in Iron the general accentual rules place the accent upon either the first or the second syllable of the accentual complex, depending upon the strength or weakness of the initial syllable's vowel. In the case of accent complexes containing a substantive, however, there is a systematic shift of the accent from the second syllable to the first when the substantive is definite. This means that, in complexes whose first syllable contains a weak vowel (*æ* or *ɪ*), there is a regular distinction between second-syllable accent (*kʰrój* 'a mill', *sírʰ tírɪsə* 'a red flag') and first-syllable accent (*kʰrój* 'the mill', *sírʰ tírɪsə* 'the red flag').

The historical explanation for this accent shift may be seen in the modern Digor definite article *i* (e.g., *i kizgæ* 'the girl', *i bæx* 'the horse'—cf. Isaev 1966: 33–34). Early Iron, too, evidently had such an article, which, when counted as the first syllable of its accent complex, caused the accent to be placed one syllable to the left (**kúrój(n)-* vs. **i kúrój(n)-*), just as other prefixed monosyllabic elements containing a weak vowel do (cf. *mæ kʰrój* 'my mill'). When the **i* was lost—as initial **i* was lost elsewhere in Iron (cf. *mísɪn* 'compose' vs. Digor *imɪsɪn*, *ssæʒ* 'twenty' vs. Digor *insæj*)—the shifted accent remained grammaticalized as an indication of definiteness.²⁷

27. I suspect that another trace of the article **i* in Iron is to be found in the two terms for 'other,' *innæ* and *annæ*, both of which are regarded as reflexes of Iranian **anya-*. *Innæ* is used specifically to refer to the second of a series; it is likely that it originally reflected the 'other' of a set of two (*alter*). Being post facto determined (as 'the other' of a specified pair), it could have been regularly associated with the definite article as **i annæ*, and could thus ultimately have merged into a unit *innæ*. Subsequently its use was evidently expanded from '*the other of two'—cf., e.g., Digor *ci bazurdtaɪ ibæl ærzajuɪ*, *wonæj jew bazur xussægi bazur æj*, *innæ bazur ba iɣælzægi bazur æj* 'of the wings that grow on it [the bat], one wing is the wing of sleep, the other (*innæ*) wing is the wing of wakefulness' (Miller 1902: 61)—to 'second of a series'—cf. *kæstærtæj iw ... bæxti ... abasta*; *innæ ta sin wazægdonmæ sæ nimattæ cird bask'æfta*; "*midæmæ ut næ wazʒitæ*" *zayta sin wæd annæ kæstær* 'one (*iw*) of the younger ones ... tethered ... the horses; another (*innæ*) quickly brought their felt boots into the guestroom for them; "Come in, our guests," another (*annæ*) younger one then said to them'; *wærtæ Rac'mæ—sæ iwti fænd*, *wæd se 'nnæti*—*Tægiatæm*, *annæti zond—Kæsxæʒi 'rdæm* 'the intention of some of them (*sæ*

Bibliography

- Abaev, V. I. 1939. "Ritmika osetinskoj reči (Satzakzent)." In *Iz osetinskogo èposa*, pp. 96–134. Moscow & Leningrad: Iz-vo Akademii nauk SSSR.
- . 1949. *Osetinskij jazyk i fol'klor*. Moscow & Leningrad: Iz-vo Akademii nauk SSSR.
- . 1952. "Grammaticeskij očerk osetinskogo jazyka." in A. M. Kasaev, *Osetinsko-russkij slovar'*, pp. 441–539. Moscow: Gosudarstvennoe iz-vo inostrannyx i nacional'nyx slovarej.
- . 1958–89. *Istoriko-ètimologičeskij slovar' osetinskogo jazyka*. Vol. 1 (1958), Moscow & Leningrad: Iz-vo AN SSSR; Vols. 2 (1973), 3 (1979), 4 (1989), Leningrad: Iz-vo Nauka.
- . 1964. *A Grammatical Sketch of Ossetic*, ed. Herbert H. Paper, trans. Steven P. Hill (Indiana University Research Center in Anthropology, Folklore, and Linguistics Publication 35). Bloomington: Indiana University.
- . 1968. "O perekrestnyx izoglossax." *Ètimologija* 1966: 247–63.
- Axvlediani, G. S. 1924. "Tri t v osetinskom i ix otnošenje k sootvetstvujuščim dr. iranskim zvukam." *Jahrbuch der georgischen sprachwissenschaftlichen Gesellschaft* 1–2: 22–31.
- , ed. 1963. *Grammatika osetinskogo jazyka*, vol. 1. Oržonikize, North Ossetia (Russian Federation): Naučno-issledovatel'skij institut pri Sovete ministrov Severo-osetinskoj ASSR.
- Bagaev, N. K. 1965. *Sovremennij osetinskij jazyk*, Part 1. Oržonikize, North Ossetia (Russian Federation): Severo-osetinskoe knižnoe iz-vo.
- Benveniste, E. 1959. *Études sur la langue ossète*. Paris: Klincksieck.
- Gershevitch, Ilya. 1985. *Philologia Iranica*, ed. Nicholas Sims-Williams. Wiesbaden: Reichert.
- Henderson, Eugénie J. A. 1949. "A Phonetic Study of Western Ossetic." *Bulletin of the School of Oriental and African Studies* 13: 36–79.
- Isaev, M. I. 1954. "O vokalizme osetinskogo jazyka." *Trudy Instituta jazykoznanija AN SSSR* 3: 227–49.
- . 1966. *Digorskij dialekt osetinskogo jazyka*. Moscow: Izdatel'stvo Nauka.
- . 1987. "Osetinskij." In *Osnovy iranskogo jazykoznanija. Novoiran-skie jazyki: Vostočnaja gruppa*, ed. V. S. Rastorgueva, pp. 537–643. Moscow: Nauka.

iwti) is (to go) to Rac', (that) of others of them (*se 'nnæti* ← *sæ innæti*)—to the Tagaurs, the thought of others (*annæti*)—(is) toward Kabardia' (Abaev 1958 1: 54).

- Job, D. M. 1982. "Ossetic Verb Stem Morphophonology." In *Papers from the Second Conference on the Non-Slavic Languages of the USSR*, *Folia Slavica* 5: 214–23.
- Karažaev, Ju. D. 1984. "Zametki po osetinkoj fonologii." *Problemy osetinskogo jazykoznanija* 1: 85–105.
- Kosven, M. O.; L. I. Lavrov; G. A. Nersesov; and X. O. Xašaev, eds. 1960. "Osetiny." In *Narody Kavkaza*, vol. 1. Moscow: Izdatel'stvo Akademii nauk SSSR.
- Miller, Vsev. 1902. *Digorskija skazanija*. Moscow: Tipografija Varvary Gatsukъ.
- Miller, Vsevolod. 1962. *Jazyk osetin*, trans. M. I. Isaev. Moscow & Leningrad: Iz-vo Akademii nauk SSSR (Russian original 1910).
- Sokolova, V. S. 1953. *Očerki po fonetike iranskix jazykov*, vol. 2. Moscow & Leningrad: Izdatel'stvo Akademii nauk SSSR.
- Thordarson, Fridrik. 1989. "Ossetic." in *Compendium Linguarum Iranicarum*, ed. Rüdiger Schmitt, pp. 456–79. Wiesbaden: Reichert.
- Tybylty, A. 1929. "Xussar Irystony dialekttaë." In *Xussar Iron Adæmy Uacmystæ*, vol. 2, pp. 2–8. Cxinval, South Ossetia (Georgia): Xussar Irystony Bæstæzonæg Axuyr-Irtasæg Institut.
- Zgusta, Ladislav. 1955. *Die Personennamen griechischer Städte der nördlichen Schwarzmeerküste*. Prague: Československá Akademie Věd.
- . 1987. *The Old Ossetic Inscription from the River Zelenčuk*. Vienna: Verlag der österreichischen Akademie der Wissenschaften.

This page intentionally left blank.

Pashto Phonology

Josef Elfenbein

Johannes Gutenberg–Universität Mainz

36.1. Introduction

Pashto is one of the two official languages of Afghanistan, having been accorded its status in 1936 beside Dari Persian; since 1979 five other languages have been recognized as “national” languages of Afghanistan as well. There have also always been as many or more Pashto speakers in Pakistan as in Afghanistan, the majority of them living in Northwest Frontier Province (NWFP) and in Baluchistan. In the extremely disturbed conditions in the area since the Soviet occupation of Afghanistan in 1979, there have been many millions of Pashto-speaking refugees from Afghanistan who have fled to Pakistan (and to Iran), and it is now (1993) quite impossible to provide meaningful statistics regarding either numbers of speakers or their present whereabouts. In this article, therefore, everything geographical must be referred to the pre-1979 situation, as far as Afghanistan is concerned.

In 1979 Pashto was, and is certainly now still, the most important Iranian language after Persian in terms of the number of its speakers, who totalled in 1979 at a conservative estimate at least 8 million in Afghanistan, and ca. 12 million (including refugees) in Pakistan, making a total of at least 20 million, quite likely more, even many more. Pashto is spoken as the main language throughout central-east, south, and southwest Afghanistan; in Pakistan, in north Baluchistan from Quetta throughout NWFP to northwards of Peshawar as far as Dir in Gilgit, and eastward to the Indus River in Swat and Buner, as well as in all large cities in Pakistan (there is a very large colony of perhaps 1 million Pashto speakers in Karachi). There are also largish colonies of Pashto speakers in north-central Afghanistan, settlers since the end of the 19th century, in Maimana and in Mazar-i Sharif, and in

AUTHOR'S NOTE: Abbreviations: Ac, Acakzai; Afr, Afṛīdī; Ar, Arabic; Bg, Bangaš; Bn, Ban-nūčī; Gh, Ghilzay; IA, Indo-Aryan; Ir, (Old) Iranian; Kand, Kandahārī; Lhd, Lahndā; Moh, Mohmand; NP, (New) Persian; NWFP, Northwest Frontier Province; Or, Orakzay; Pesh, Peshāwār(i); Pnj, Panjābī; Psht, Pashto; Sher, Sherānī; Si, Sindhī; Tar, Tarīn; Wan, Wanetsi, Waṇecī; Waz, Wazīrī; Yus, Yūsufzai.

a few other places farther east. The number of Pashto speakers in Kabul is at present unknown, due to massive numbers of refugees (see Penzl 1955, 1–4).

There are also not insignificant colonies of Pashto-speakers (perhaps 250,000+) in India (mainly in Uttar Pradesh), in the Gulf States, on the south Persian coast (trading colonies e.g. in Bandar Abbas and Minab), on the eastern Persian border, as well as in the larger settlements of Afghani Sistan—and in several cities in the north of England.

Pashto is a major subject of study and instruction in the University of Kabul as well as in the University of Peshawar, and there are active Academies in both cities which publish in the language—as well as other publishers. It is also taught in schools in Afghanistan and in NWFP; in the latter it has a *de facto* status as the “provincial language.”

36.2. A note on history

Pashto is the native and now customary name for the language in Europe, the older name “Afghan” (and the even older variants “Pushtu,” “Pukhtu” etc.) having now dropped out of use; its tribal speakers are called Pashtuns (and not “Afghans”). It is not known where the original Pashtun settlement was, or when settlement in their present habitat took place. But probably significant is an old tradition that they had an ancient home in the Sulayman Mountains northeast of Quetta, called in Pashto *da kase yar*, whence the tribal name Kāṣi, and their appellation Kaṣi amongst the Ormurs of Waziristan. From there they would have spread out, perhaps rather more than 1000 years ago, to their present habitats.

36.3. The name of the language and people

The name Pashto (*Páḥto* in Pashto, see § 36.8), both for the language and for tribal law, and the name Pashtun (*Páḥtun*) for the people are employed by all tribes (but cf. *paḥtó* ‘the language’, *paḥtyó* ‘sense of honor’ in the Andar tribe, east of Gardez), and cannot have anything to do with the Paktyes, a people mentioned in Herodotus. It is now agreed by all scholars that the names, first known from the 16th century, go back respectively to **parsawā* ‘the Pashto language’ and to **parswāna*- ‘the Pashtuns’ (first adumbrated by Markwart in 1896, finally established by Morgenstierne 1940: 143–44). These names may connect with the *Parsietai* and *Parsioi* mentioned in

Ptolemy, said to be a people inhabiting the Paropamisus mountains. Not excluded is an ultimate connection with the *Parśu-*, a warrior tribe of north-western India, mentioned in Pāṇini (before the 4th century B.C.), and perhaps even with the name *Parsa-*. Indo-European ethnic names often spread over wide areas and are used by widely scattered different peoples. The Indian name Paṭhān, first known from the 16th century but certainly much older, comes from an early Pashto form **paṣtān*.

Some Pashtun tribal names are also undoubtedly old: e.g. the Afridis are most probably to be identified with Herodotus' Aparytai, and Ptolemy's Paroutai (if they are the same people) and the Avestan Parutas (cf. Gershevitch 1959: 174–75).

36.4. The name “Afghan”

The Persian name Afghan (*Afγān*), of disputed origin, is of much older occurrence, first in the 10th-century Persian geography *Hudūd al-‘Ālam*, and may be related to the name Avagāṇa, a people mentioned in a 6th-century Indian work (see Morgenstierne 1960, 1983a; and an interesting possibility suggested by Vercellin 1983).

36.5. Pashto literature

After Persian, Pashto has the oldest literary tradition amongst Modern Iranian languages, with an authentic literary culture going back to the 16th century. The earliest datable manuscript is dated 1651, and contains the *Xayr ul-bayān* (see MacKenzie 1964), the religious polemics of Bāyazīd An-sārī, who was active a century earlier (d. 1585). Perhaps the most famous poet in Pashto is Xūshāl Xān Xaṭak (1613–1689), whose diwān is still daily fare. Other classic writers are ‘Abd al-Raḥmān (“Bābā”), ‘Abd al-Hamīd (both fl. ca. 1700; see MacKenzie 1965; Morgenstierne 1960: 220–21, § iii, “Pashto literature”; Darmesteter 1888–90; general bibliographies are to be found in Skjærvø 1989: 409–10; Morgenstierne 1960; see also Grierson 1921: 14–16, for older literature). Publishing is very active in both Kabul and Peshawar, to some extent rival centers, Peshawar “hard” in a modified Urdu script, Kabul “soft” in a newly adapted Pashto script (see § 36.8.1 for the terms “hard” and “soft,” and §§ 36.10, 36.15 for scripts).

Concerning the much-discussed manuscript of the *Pəṭa Xazāna* (“Secret Treasury”), first brought to light by Prof. ‘Abd al-Ḥayy Ḥabībī (“Abdul Hay

Habibi”) and published in Kabul in 1944, supposed to contain Pashto poetry going back to the 8th century, nothing which has been published has been sufficient to remove the grave doubts expressed by all specialists about its authenticity (see Morgenstierne 1960: 220; personal communications from Prof. Morgenstierne; Penzl 1955: 4; latest discussion in Loy 1987). My own opinion, based also upon discussions amongst Pashtuns in Quetta, is that the manuscript is a pious fraud, compiled in the Loralai district of Pakistani Baluchistan, probably in the late 18th century.

36.6. The Pashto language

Pashto is an Iranian language of quite specific northeastern character, with closest relationships to the Saka group of Middle Iranian; but no known Middle Iranian language can claim to be its ancestor (see Skjærvø 1989 § 4.2.1.1.5.1; MacKenzie 1987 § 2). Pashto dialects are as ramified as the tribal structures underlying them, of which they are at once an expression and a badge of identity. Indeed, it has been well put (Barth 1964) that “Pathan tribal structures are primarily linguistic [structures], needing constant defense by argument ... as well as by other means, from other Pathans.” Dialects within each of the traditional dialect groups are more or less mutually intelligible, but mutual intelligibility between different groups is much less, with Waziri especially so; and one “dialect,” Wanetsi, which has some claim to separate language status, is quite unintelligible to other Pashto speakers.

Dialect differences lie primarily in phonology and lexicon: the morphology and syntax are, again with the exception of Wanetsi, quite remarkably uniform.

Loanwords have been traced in Pashto as far back as the third century B.C., and include words from Greek and probably Old Persian. Since Islamic times there is to be found the usual lot of loanwords from Persian and Arabic, as well as many from Indo-Aryan (mainly from Sindhi and Lahndā), and latterly since the end of the 18th century an increasingly large number of loanwords from Hindi/Urdu and English.

Amongst the modern Iranian languages the phonology of Pashto is of middle complexity, lying between the simplest—such as Persian or Balochi—and the most complex, such as the Pamir languages. Its morphology, however, is very complex. The following three phonological characteristics leap to the eye (or ear) as typical of Pashto.

36.6.1. Consonant clusters

A liking for (or a lack of aversion to) word-initial consonant clusters, in all dialects; some hundred such clusters occur, e.g. two consonants: *cṣəl*, *psar-láy*, *ṣkanjəl*, *nṣəl*, *nmunj*, *nmasáy*, *jyastəl*, *lmar*; three consonants are not at all unusual either: *swaxtáy*, *xwrinedəl*, *xwla*, *nywaḡ-*, etc. There are no initial four-consonant clusters in standard Pashto. This phenomenon has given rise to the well-known story comparing Pashto unfavorably (“the bray of an ass”) to the other major Islamic languages Arabic, Persian, Turkish, or Urdu. These initial consonant clusters are much less in evidence in Wanetsi. Consonant gemination, however, is unknown in Pashto.

36.6.2. Phonological instability

These initial consonant clusters are rather unstable, and give rise to the characteristic Pashto metatheses and dissimilations, not only as between different dialects, but also within the same dialect, e.g. *byarg*, *ybarg*; *lwan-*, *nwal-*, etc. see §§ 36.16.3, 36.16.4.

36.6.3. Stress

Pashto has phonemic variable stress, unique amongst Modern Iranian languages. For example: the pronouns

	<i>haṣá</i>	‘that, he’	Wan <i>aṣá</i>
	<i>daṣá</i>	‘this, he’	Wan <i>daṣá</i>
vs.	<i>háya</i>	‘that one, he’	
		(emphatic)	Wan <i>áya</i>
	<i>dáya</i>	‘this one, he’	
		(emphatic)	Wan <i>dáya</i>

are also differently inflected. Further examples are:

<i>gorá</i>	‘European’ vs.	<i>góra</i>	‘look!’
<i>palítá</i>	‘wick’	<i>palíta</i>	‘whore’
<i>wārṣ</i>	‘small (m.pl.)’	<i>wáṛṣ</i>	‘all (m. sg.)’
<i>láta</i>	‘district’	<i>latá</i>	‘a kick’
<i>áspa</i>	‘mare’	<i>aspá</i>	‘typhus’
<i>térə</i>	‘last, past’	<i>terṣ</i>	‘sharp’
Kak <i>wažá</i>	‘shoulder’	Kak <i>wáža</i>	‘garlic’

Stress is used in verbal forms to distinguish e.g. aspect:

<i>kṣenāstələm</i>	‘I was sitting’ vs.	<i>kṣénāstələm</i>	‘I sat down’
<i>bakṣenám</i>	‘I shall be sitting’ vs.	<i>bakṣénəm</i>	‘I shall sit’

as well as mood:

preǵdám ‘I leave’ vs. *preǵdám* ‘that I leave’

There are three degrees of stress: strong, (“loud” in Penzl 1955), medium, and weak. Stress can fall on any syllable of a word, but mainly it is confined to the first, last, or penultimate syllable. The position of stress is not indicated in writing, but good dictionaries indicate at least the position of strong stress. Examples:

<i>prewatál</i>	‘to fall, they (m.) were falling’	(strong)
<i>préwatàl</i>	‘they (m.) fell’	(strong, weak)
<i>prewátay</i>	‘fallen (m.sg.)’	(medium)

In this article only strong stress will be marked, by an acute (´) over the vowel.

36.7. Pashto dialectology

It is not possible to properly describe Pashto phonology independently of a thorough and systematic dialectology. But Pashto dialects are so ramified and their interrelationships so complex as to defy up to now any attempt at a comprehensive scientific description: one simply has not enough data of the right kind. In these circumstances I must, in what follows, perforce renounce any claim to comprehensiveness.

We have a good description of the dialect of Kandahar (Penzl 1955), and a reasonably good one of Wazīrī (Lorimer 1902). Good studies of the other dialects of Pakistan are mostly lacking: the material in Grierson 1921 gives but a smattering of the dialect situation, and is inadequate by any modern phonetic standard; but if used with care it can still be useful. The dialect description in Kieffer 1983: 504ff. and Table 10) is necessarily very sketchy, lacking sufficient detail. Wanetsi has been partly described by Morgenstierne (1930, 1973) and by Elfenbein (1967, 1984), and needs much further work. Good textual material is available in Afīdī (Morgenstierne 1931), as well as to a lesser extent in Ghilzay and Xalīlī. The dialect material in Morgenstierne 1927 is to be used with caution; much better material has been prepared by MacKenzie and Elfenbein for a second edition thereof, not yet published. There is in my possession a fair amount of somewhat unsystematic collections in the Kākaṛī and Acaḳzay dialects of Pakistani Baluchistan. The ALA collections (see Kieffer 1974) are of course confined to Afghanistan, based upon geographical point-survey questionnaires for the situation as it was before 1979, and have of course lost some of their validity since then in the post-1979 chaos. What follows is therefore far from what I should wish, based as it has had to be only on data as explained above.

36.8. Pashto dialects

There are various ways in which Pashto dialects can be classified, and none of them is entirely satisfactory because (1) dialect isoglosses cross any possible dialect boundaries; (2) geographical classification alone does not—cannot—take enough account of tribal distinctions, which are basic and essential; (3) a classification by tribe alone fails to take into enough account the essential geographical facts. The by now usual dialect division, based as it is on phonological criteria alone, uses a mixture of geography and tribal description. But large tribes, such as the Afrīdī, Ghilzay, and Kākaṛ defy placement in any single class; this is also true of the nomadic tribes.

The dialect division given below, it must be emphasized, is not visible to the naked eye in the script employed for writing Pashto, due to the adoption of a (fairly) unified written standard (see §§ 36.10, 36.15).

36.8.1. *Dialect classification*

The now generally accepted fourfold division given below is based on the pronunciation of the four consonants transcribed *ṣ*, *ǧ*, *c*, and *j*, and is mainly geographical.

36.8.1.1. *Northwest Group* (“semi-hard” group)

Includes east and northeast Afghanistan, Central Ghilzay, Jalālābād, some Afrīdī.

ṣ = [ç+] often transcribed *ṣ*, or *sh*

Amongst the Ghilzay, those west of Kabul often have *ṣ* = [ʃ], but details are lacking.

ǧ = [j+] (voiced [ç]), often transcribed *ǧ* or *ǰ*

c = [s], transcribed *s*

j = [z], transcribed *z*

36.8.1.2. *Southwest Group* (“soft” group)

Includes Kandahar (the basis of the Afghan written standard), Farāh, Herat, South Ghilzay.

ṣ = [ʃ], often transcribed *ṣ*

Amongst the South Ghilzay, the Sliman Xēl have *ṣ* = [ʃ–], strongly retracted.

ǧ = [z], often transcribed *ǧ*

c = [ts], transcribed *c* or *ts*

j = [dz], transcribed *j* or *dz*

36.8.1.3. *Northeast Group* (“hard” group)

Includes Peshāwar, Shīnwārī, northern parts of NWFP, Yūsufzai, Mohmand, northeast Ghilzay, Hazara, Bangaš, Orakzay, some Afrīdī.

$\check{x}$ = [x], transcribed *x* or *kh*

$\check{g}$ = [g], transcribed *g*

c = [s], transcribed *s*

j = [z], transcribed *z*

Amongst the Afrīdī, the subsections Malikdīn Xēl, Zaxa Xēl, and Kukī Xēl differ, but details are lacking.

36.8.1.4. *Southeast Group* (“soft” group)

Includes the southern parts of NWFP, with Sherānī, Bannū, Wazīrī; Baluchistan, with Kākaṛī, Acakzay, Tarīn (Tōr Tarīn, Spīn Tarīn, Bōr Tarīn), Wanetsi (Waṇecī).

$\check{x}$ = [ʃ or ʒ], transcribed *š* or *ž*, respectively

$\check{g}$ = [ʒ or ʒ], transcribed *ž* or *ž*, respectively

c = [ts], transcribed *c* or *ts*

j = [dz], transcribed *j* or *dz*

Wazīrī can be subdivided into north and south subdialects. Of the two pronunciations of $\check{x}$ and $\check{g}$ given above, [ʃ] and [ʒ] are more usual in south Wazīrī, north Kākaṛī, Bannūčī, and Tarīn; Waṇecī has only [ʃ] and [ʒ], respectively. Acakzay and Sherānī could be put into the southwest group. South Kākaṛī is very similar to Kandahārī.

It is perhaps to be expected that the huge Afrīdī, Kākaṛī, and Ghilzay tribes, many of whom are nomadic or semi-nomadic, share characteristics with the northeast, northwest, and southwest groups. As will be seen, there are also relationships, based on other criteria, between northeast and southwest, southeast and southwest, and northwest and northeast groups. There is no definite line between “hard” and “soft” Pashto.

36.9. The segmental consonant phonemes of Pashto

See Table 36-1.

36.9.1. *Notes to Table 36-1*

p [p] voiceless bilabial stop, somewhat aspirated prevocally, especially in a stressed syllable

b [b] voiced bilabial stop, unaspirated

Table 36-1. Pashto Consonants

	Labial	Dental	Palato- alveolar	Retroflex	Velar	Uvular	Glottal
Stops	p b	t d		ʈ ɖ	k g	(q)	(ʔ)
Affricates		c j	č ǰ				
Fricatives	(f)				x ɣ	h	
Spirants		s z	š ʒ	ʂ ʐ			
Nasals	m	n		ɳ	(ŋ)		
Lateral		l					
Flaps		r		ɽ			
Semivowels	w	y					

t [t] voiceless post-dental stop, somewhat aspirated prevocally in a stressed syllable

d [d] voiced post-dental stop, unaspirated

ʈ [ʈ] voiceless retroflex stop, aspirated somewhat prevocally in a stressed syllable; occurs only in loanwords from IA. Sometimes transcribed *ṭ* (e.g. in MacKenzie 1987)

ɖ [ɖ] voiced retroflex stop, unaspirated; occurs only in loanwords from IA. Sometimes transcribed *ḍ* (e.g. in MacKenzie 1987)

k [k] voiceless velar stop, aspirated somewhat when prevocalic in a stressed syllable

g [g] voiced velar stop, unaspirated

q [q] occurs only as an “elegance” in loanwords from Arabic; otherwise = *k*

ʔ [ʔ] occurs only as an “elegance” in loanwords from Arabic with ‘ayn or hamza; it is otherwise simply dropped, with or without lengthening the previous vowel

c [ts] see § 36.8.1. Voiceless post-dental affricate, slightly aspirated in stressed syllables; a single phoneme (despite Penzl 1955 § 22.2f.), since e.g. it forms syllable boundaries in word-final position, and can form initial clusters with two other consonants (there being no 4-consonant initial clusters in Standard Pashto): *cṭmæx* ‘lining’, *cɣəl* (m.), *cɣla* (f.), pret. ‘filled, packed’; *cṛakəhāṛ*, Kak *cṛkəhāṛ* ‘dripping’, etc.; and it metathesises as *c*, e.g. *cnəy*, *ncəy* ‘woolen thread’

j [dz] see § 36.8.1. Voiced post-dental affricate, a single phoneme for the same reasons as under *c*. Not aspirated

č [tʃ] voiceless palato-alveolar affricate, a single phoneme for the same reasons as under *c*. Slightly aspirated in stressed syllables

ǰ [dʒ] voiced palato-alveolar affricate, a single phoneme for the same reasons as under *c*. Unaspirated

f [f] voiceless labio-dental fricative, occurs only as an “elegance” in loan-words, mainly from Persian or Arabic, but also from English; otherwise = *p*. (see § 36.17.4)

x [x] voiceless velar fricative

ɣ [ɣ] voiced velar fricative

h [h] voiceless glottal fricative, dropped in most southwest and some southeast dialects, e.g. in south Kak, where it is an “elegance”; also dropped in Wan.

s [s] voiceless post-dental spirant

z [z] voiced post-dental spirant

š [ʃ] voiceless palato-alveolar spirant

ž [ʒ] voiced palato-alveolar spirant

ǰ see § 36.8.1. In dialects of the southwest group and parts of the southeast, a voiceless retroflex spirant

ǧ see § 36.8.1. In dialects of the southwest group and parts of the southeast, a voiced retroflex spirant

m [m] voiced bilabial nasal

n [n] voiced post-dental nasal; non-phonemic allophones are [ŋ] before *k* and *g*, [ɳ] before *t* and *ɖ*, and [ɲ] before *č* and *ǰ*

ɳ [ɳ] voiced retroflex nasal flap; cannot stand in initial position; sometimes transcribed *ñ*

l [l] voiced post-dental lateral, non-fricative, medium clear

r [r] voiced post-dental flap, single in most dialects

ɽ [ɽ] voiced back-alveolar retroflex flap; sometimes transcribed *ṛ*

w [w] voiced labio-dental continuant

y [j] voiced palatal continuant

36.9.2. *Minimal pairs*

A few minimal pairs follow for *š* : *ǰ*, *ǧ* : *ž*, *r* : *ɽ*, *ɳ* : *n*. (The translations, here and in what follows, are “minimum translations,” and serve only to identify the word.)

šāxá : *šāxá* ‘branch : narrow path’; *špa* : *špa* ‘night : leg’
žay : *žay* ‘yoke : cry’; *žay* : *žay* ‘edge : sack tie-rope’
rap : *rap* ‘instant : friable soil’; *rang* : *rang* ‘color : destructive’
run : *run* ‘bright : thigh’; *maná* : *maná* ‘believe! : apple’

Decisive, however, is the history of the sounds. For example *š* < **-ršt-*, **-štr-*; *-št-* < **-rsw-*, **-rsy-*; *š* is a usual development of *š* in early loanwords from NP (*peš*, *dušmān*, etc.; see § 36.15.1); whereas *š* < e.g. **-xš-*, **sy-*; and *g* < e.g. **-čiy-*, **-ržn-*, *-rš-*, *-rž-*; but *ž* < **j*, **ziy*; further, *r* < e.g. **rt*, **rθ*, **rd*, **žd*; but *r* < e.g. **r*, **-θr-*; *ŋ* < e.g. **xšn*, **rn*; but *n* < e.g. **n*, **xn*, **θn*, **fn*. I give no examples, as a very good summary of the history of Pashto sounds is to be found in Skjaervø 1989 § 4.2.1.7. Much of this material comes from Morgenstierne 1927; a projected second edition (from Morgenstierne’s Nachlass, by MacKenzie and Elfenbein) is as yet unpublished.

36.10. Remarks on the Pashto script for consonants

(For vowels and diphthongs see § 36.15.)

As remarked above (§ 36.8), the unified script conceals dialect differences. The Persian script is used, with the following additions:

š = *ش*, *g* = *ځ*, *c* = *خ*, *j* = *ج* (but older and Peshawar *خ*), *t* = *ټ*, *d* = *ډ*, *r* = *ړ*, *ŋ* = *ښ* (but older *نږ*); *g* is written *گ* or *گی*. The letters *ث*, *ص*, *ض*, *ط*, *ظ*, *ز*, *ح*, *ع*, *ق* are used only in loanwords from Arabic, and are pronounced as in Persian (for the letters ‘*ayn*, *q*, and *f*’, see § 36.9). They can be transcribed as *s*, *z*, *z*, *t*, *ž*, *h*, ‘, *q* respectively, to show their origin, when desired (other systems are also in use). In Peshawar printers use *ٹ* *ڈ* *ڑ* *ڻ* for *t* *d* *r* *ŋ* respectively, as in Urdu.

36.10.1. History of the script

The special Pashto letters were probably developed in Peshawar, before the 16th century. In the earliest known Pashto manuscript with a definite date, namely 1651 (see MacKenzie 1959: 233), *j* is written *ځ*, *ž* is written *ځ*, and *g* is written *ځ*.

These signs were abandoned shortly after 1690. The system presently used in Afghanistan (with *j* = *خ* for older *خ* [thus distinguishing *j* and *c*], as well as extra vowel signs, see § 36.15), was only officially adopted in 1958.

36.11. Remarks on consonants in the dialects

In what follows, “Psht” means ‘standard Pashto’.

36.11.1. *Southwest dialects*

Kand depalatalises all forms of the verb *šwəl* ‘become’ (> Kand *swəl*); cf. also Kand *spǝǵa* ‘louse’ (< *špǝǵa*). Depalatalisation of *š* to *s* is also found in south Gh (*sǝga* ‘sand’ < *šǝga*). The postposition *kǝ* ‘in’ is usually *ke*, as in most dialects. *h* is only pronounced as an “elegance.”

36.11.2. *Southeast dialects*

Depalatalization (see also § 36.16.1) is marked in some cases: *š* > *s* very commonly in Kak, Wan, Tar: in all three we have *sǝga* ‘sand’, *sǝ* ‘becomes’ (< *šǝ*), stem *sw-* < *šw-*. In Kak, Tar also *z* < *ž* in some words. There is also a marked avoidance of *f*, *h*, and *ǰ*, which become, respectively, *p*, *ø*, and *j*. In Kak we have, with *s* for *š*, *ysǝy* ‘arrow’; *kasr* ‘younger (brother)’; *lwas-* ‘to milk’; *sta* ‘is’; *kǝ* ‘in’ is always *ke*. *h* is heard only as an “elegance” in south Kak. On the other hand we have palatalization in Kak *wǝlyǎr* ‘standing’ (< *walǎr*).

The sounds *š* and *ǵ* receive variable treatment: Kak has both *š* [s] and *š* [ʃ], but *š* is commoner; both *ž* [z] and *ž*, but *ž* is commoner: e.g. *žyǎnzy/ga* ‘flea’, cf. *ǵyǎnja*; *tož-* ‘to shave’ cf. *toǵ-*; *wǎžay* ‘ear of corn’, cf. *wǎǵay*. But we have also e.g. *pišǝ* ‘cat’, cf. Psht *pišǝ* (see § 36.17.2). Some Kak dialects go even further with *žǎwla* ‘resin’, *žawǝr* ‘deep’, for Psht *žǎwla*, *žawǝr*; details are lacking. For Wan, *žyāl* ‘gravel’ for Psht *ǵyal* is typical of the change *ǵ* > *ž*.

36.11.3. *Northwest dialects*

Dearth of information makes it possible only to note: Central Gh has *c* > *č*, and *j* > *ǰ*, e.g. *čexǎ* ‘near’; stem *ǰ-* ‘go’. Central Gh is reported to have a two-fold pronunciation of *š* and *ǵ*: some groups have *š* with *ǰ*, whereas other groups have *š* with *g* and, as has been noted by e.g. Kieffer (1983: 505, “B” dialects) stands as a transitional dialect between northeast and southwest types.

36.11.4. *Northeast dialects*

The northeast group presents perhaps the most complicated picture of all, with each tribe and district possessing its own unique dialect; unfortunately precise details are still lacking. In Yus, Pesh, and Moh *ž* > *ǰ*; this is often the case in Afr as well, where *ž* > *z* also in some cases, e.g. Psht *wrúja* ‘eyebrows’ > Afr *wrúzya*; Psht *wǎža* ‘a tendon in the neck’ > Afr *wǎža*; Psht *žǝr* ‘quick’

> Afr *zər*. Northeast Gh tends to palatalize all consonants standing before front vowels: e.g. *ǰǰja* ‘woman’ has pl. *ǰǰje*; *sǎŋga* ‘branch’ has pl. *sǎŋje*. Moh often has *š* for *ǰ*, contrary to “hard” group characteristics. Yus, Moh, and Afr have clearly been influenced in their consonantism by IA; as has been noted (e.g. by MacKenzie 1987: 551) the whole consonant system of the northeast group has tended for several hundred years—likely for much longer—to lose sounds which are foreign to IA, with the exception of *x*, *ɣ*, and *z*. In Peshawar one hears occasionally *c* as [ts] and *j* as [dz], but whether these are mere “elegances” or historic survivals is hard to say. Nowadays the two sounds verge respectively on [t+] and [d+], approximating to IA; but there are no phonemic aspirates.

In the central northeast area, where there is extensive multilingualism with IA languages (e.g. Lahndā, Hindkō, Panjābi), *ŋ* is pronounced identically to Lhd [ŋr] (cf. the older writing *ڼ* for *ځ*, still used in Peshawar, see § 36.10). Further, one hears *Ųŋ > Ųr*, and *Ųŋj > Ųj*, as in Lhd.

36.12. Other dialect features

It has been noted (by Morgenstierne) that the syntax of the Psht of this area has also been profoundly influenced by e.g. Lhd, in that whole sentences from northeast Psht can be put word-for-word into Lhd, and vice versa. There is also, as in Lhd, a spontaneous change of *ŋ* to *Ųr*, e.g. *rũŲr* ‘lighted’ < *rũŋr* < **rũxšna*-.

Examples of palatalization are Afr *máinde*, northeast Gh *myénde*, Moh *myǎnde* ‘mothers’; see § 36.16.1.1.

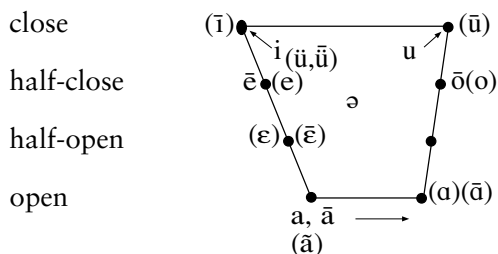
It is in the northeast area that the origins of Psht literature are to be sought. When the alphabet was fixed in the 16th century, the present sound changes of the area had not yet taken place, and the alphabet then fitted the northeast Psht sound system much more closely than it now does: e.g. the oppositions *x* : *ǰ* and *g* : *ǰ* still existed. The velarization of consonants (the “hardening” process) most probably began amongst the Yusufzais and Mohmands at some time in the 16th century, and spread out from them. (But vowel changes are different, see § 36.13.7). It was Morgenstierne who first suggested this history (see Morgenstierne 1932: 17), which still seems today much the most likely (against Penzl 1955 § 4.4, who thought of a cradle in the southwest because of the present-day precise fit of phonology and script there, and only there; for a much more complete discussion of these matters see MacKenzie 1959; Kieffer 1974: 28–29.)

36.13. Stressed vowels

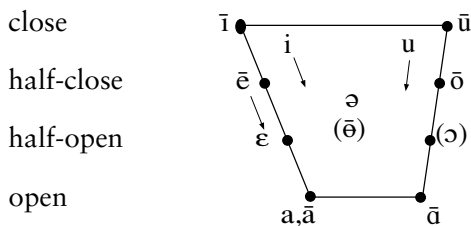
Vowels in brackets are of uncertain phonemic status. Macrons are used instead of colons for long vowels, and cardinal positions are indicated by dots.

$\ddot{u} = [y]$, $i = [ɪ]$, $u = [ʊ]$, $\ddot{a} = [æ]$; arrows indicate regions of variation.

36.13.1. Southwest/Southeast

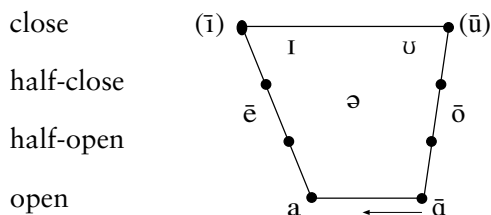


36.13.2. Northwest/Northeast



36.13.3. Kandahārī

It is also useful to indicate the much simpler vowel system of southwest standard Kandahārī, well-described by Penzl (1955); it is entirely symmetrical:



It is usual to omit macrons on i , u , e , and o in descriptions of Psht based on Kand, since short and long $[e]$ and $[o]$ do not contrast in stressed position; for i and u , see below.

36.13.4. *Southwest vowels*

In Kand, final *-e* > *-i*, and *-o* > *-u* when unstressed, except in certain endings to preserve distinctions; e.g. pres.2sg. *-e* contrasts with 3sg. *-i*. In most dialects, and certainly in Kand and the southwest generally, the sounds *ī* and *ū* exist only as “elegances,” being commonly reduced to coincide with *i* [ɪ or i] and *u* [ʊ or u] respectively: [i, u] in final position. Like *e* and *o*, their length is not phonemically significant, but depends only on position and stress.

Before a front vowel, Kand raises stressed *é* to *ī́*, and *ó* to *ū́* (e.g. *tére* > *tī́ri*, *nóre* > *nū́ri*).

The only subtle point concerns the status of *ə*, called “zwarakay” by Psht grammarians (the best discussion is in MacKenzie 1959: 234ff.). In unstressed position there is no phonemic difference between *a* and *ə*, either alone or in the diphthongs *-ay* and *-əy* (see § 36.14.1). When unstressed, *ə* can also represent any short vowel, and vice versa *ə* can appear as *i* or *u*, depending on the phonetic environment (e.g. *žəbə*, *žibi*, *žubu*, in a sort of vowel harmony, see Skjærvø 1989 § 4.2.1.3.1). In stressed position, however, *ə* and *a* are phonemically distinct in all dialects:

<i>lāray</i> : <i>lōray</i>	‘defeat : bowels’
<i>mroṛá</i> : <i>mroṛə</i>	‘greased cloth : torment’
<i>srə</i> : <i>sra</i>	m.dir.pl. : f.dir.sg ‘red’ (m.dir.sg. <i>sur</i>)
<i>uḡdə</i> : <i>uḡdá</i>	m.dir.pl. : f.dir.sg. ‘long’ (m.dir.sg. <i>uḡd</i>)
<i>ṣə</i> : <i>ṣa</i>	m.dir.sg. : f.dir.sg. ‘good’
<i>ṭánda</i> : <i>ṭənda</i>	‘meeting : forehead’
<i>ḡiráy</i> : <i>ḡirəy</i>	‘elbow : maize’
<i>karáy</i> : <i>karəy</i>	‘freight : ringlet’
<i>bal</i> : <i>bəl</i>	‘lighted : other’
<i>hayá</i> : <i>hayə</i>	m.dir.sg. : m.obl.sg. ‘that, he’

36.13.5. *Southeast vowels*

As examples of sound changes in Kak, the following may serve as examples:

Psht		Kak
<i>óṣka</i>	‘teardrop’	<i>wóṣka</i> (<i>wo-</i> < <i>o-</i> is very common in Kak)
<i>lewə</i>	‘wolf’	<i>lawə</i>
<i>lewár</i>	‘husband’s brother’	<i>lawár</i>
<i>melmá</i>	‘guest’	<i>walmə</i>
<i>siḡá</i>	‘mountain sheep’	<i>šažá</i>

(w)lešt 'large span, between thumb and
little finger' (w)last

wrandár 'brother's wife' wrendiár

In Acakzay, stressed -é > -yé has been recorded in the feminine plurals

Psht

Ac

paxé 'ripe'

paxyé

tawdé 'hot'

tawdyé, etc.

It is likely that more generally most stressed é > yé in Ac, but examples are lacking.

36.13.5.1. The vowels of Wazirī

The chief characteristics are these.

36.13.5.1.1. ā

Stressed or unstressed ā > [ɔ:] in N Waz; > [ɒ:] in S Waz, written in the literature ō or o; examples are:

Psht

Waz

cār- 'to spy'

cor- 'to stalk'

māstə 'curds'

mostə

pāṇa 'leaf'

pōṇyē

wāzə 'fathom'

wōzáy

36.13.5.1.2. o

Stressed o > [œ:] or [ɛ:], written in the literature ö, ō or ē, ē; the differences are without doubt dialect-specific, but sources presently available do not give consistent details. Examples are:

Psht

Waz

boğ 'collected'

bež

yólay 'courtyard'

yōlay 'house rent'; also yēlay

koğ 'bent'

kōž; also kēž

lor 'sickle'

lör

maróxa 'married woman' māróša 'adulterous woman'

os 'now'

(w)ēs

orə 'flour'

ērə 'dough'

oğá 'shoulder'

(y)ežá

toy 'spilt'

tōē, tē; tī has also been

recorded

wrost 'rotten'

wrōst

xoğ 'sweet'

xōž

xol 'helmet'

xēl

Very commonly *o-* > *(y)e-*, as in *(y)ežá*, above; or *o-* > *(w)e-*, as in *(w)ēs*, above; *-óy* > *-ōy*, *-í*, or *-ē*, as in *toy*, above, and in *warbóy* ‘house-door frame’, Waz *warbōy*.

36.13.5.1.3. STRESSED *u* > *ī*

Psht		Waz
<i>lur</i>	‘daughter’	<i>līr</i>
<i>um</i>	‘raw’	<i>(y)īm</i>
<i>num</i>	‘name’	<i>nīm</i>
<i>lund</i>	‘wet’	<i>līmd</i>
<i>muǵ</i>	‘we’	<i>mīž</i>

Initial *u-* can become *wī-*, *(y)ī-*, or *we-*: e.g. Psht *úǵa* ‘garlic’ > Waz *wīža*; *um* ‘raw’ > *(y)īm*; *udó* ‘asleep’ > *wewd*.

Bannučī seems to go mainly with N Wazīrī.

36.13.5.1.4. MISCELLANEOUS WAZĪRĪ FORMS

The following Waz forms have also been recorded, and are given here as examples; it is at present not possible to be more specific:

Psht *ān* > Waz *on*: Psht *cānga* ‘branch’, Waz *cóngā*

Psht *a/ā* > Waz *ē*: Psht *ra/āšá* ‘heap of winnowed grain’, Waz *rēšá*

Psht *wālá* ‘irrigation channel’, Waz *wēlá*

Psht *wāzó* ‘fathom’, Waz *wō/ō/ē/īzáy*

Psht *á* > Waz *é*: Psht *špag* ‘six’, Waz *špež*

Psht *ɣwaǵ* ‘ear’, Waz *ɣ(w)é/ōžay* ‘corner’

Psht *ā* > Waz *ō*: *wōzáy*, cf. *wāzó* above

Psht *ā* > Waz *ī*: *wīzáy*, cf. *wīzó*, above

We find *-ā* in e.g. *rēšá*, above; occasionally Psht *e* = Waz *ā*: Psht *teró* = Waz *tārá*; cf. also Psht *é* palatalized as Waz *yé*, e.g. in *yyéla* f.sg. ‘flock’ vs. Psht *yéle* f.pl.

In some cases Psht *a* > Waz *ā*, cf. *mařóxa*, above (§ 36.13.5.1.2.)

36.13.5.2. *Waṇecī* (*Wanetsi*)

The vowel phonemes of Wan have not yet been satisfactorily described (but cf. Morgenstierne 1930, 1973; Elfenbein 1967, 1984) and need further investigation. Certain, however, are the phonemes *a*, *ā*, and *ə* which, as in Psht, are all distinct in stressed position. The other vowels have been recorded in a regrettable mixture of phonetic and phonemic transcriptions as *i/ī*, *u/ū*, *e/ē*, *o/ō*. It seems likely that the following statements are basically correct.

36.13.5.2.1. LONG VOWELS

The “long” vowels *ī*, *ū*, *ē*, *ō* are merely stressed “short” vowels *i*, *u*, *e*, *o* respectively.

36.13.5.2.2. *i/u*

ī and *ū* are not phonemically distinct from *i* and *u* respectively, and are pronounced [i] and [u] respectively when unstressed (and not [ɪ] and [ʊ] as in Kak), and [i:] and [u:] when stressed.

36.13.5.2.3. *e/o*

Similarly, *ē* and *e* do not contrast, nor do *ō* and *o*. Stressed *é* is transcribed *ē*, stressed *ó* transcribed *ō*. Stressed *é* is [e], stressed *ó* is [o]; unstressed *e* is [ɛ], unstressed *o* is [ɔ].

36.13.5.2.4. *é/i/ó/ú*

In stressed position, *é* contrasts with *í*; and *ó* with *ú*. Thus the contrast in 1pl. : 2pl. verbal endings *-ú* : *-ó* (cf. Elfenbein 1984: 61) is expected, whereas the pronoun oblique plurals *áyū*, *áyō* (ibid. 59) do not contrast; but the differently stressed counterparts *ayú* etc. have no forms in *-ó*. The noun oblique plurals in unstressed final *-ū* (ibid. 58) which correspond to Psht unstressed *-o* are anomalous, and the expected alternate *-o* does not occur; the pronunciation is [u:].

36.13.5.2.5. CERTAIN NOUN CLASSES

The Psht noun class of feminines in stressed *-əy*, *-ī* corresponds to the Wan class of feminines in stressed *-ī* (cf. Elfenbein 1984: 58, misprinted “Pashto stressed *-ay* class”). The pl. of *mor* ‘mother’ is in Wan either *méndī* [méndi] or *myéndē* [mjénde].

36.13.6. *Northwest vowels* (see *vowel quadrilateral* § 36.13.2)

For this little-studied group, the only certain recorded characteristic word is Central Gh *xwolyé* ‘sweat’ for Psht *xwalá*.

36.13.7. *Northeast vowels*

In stark contrast to the case in the southwest, *ī* and *i* are separate phonemes, as are *ū* and *u*, archaisms lost elsewhere. On the other hand, we find great vowel changes taking place in the east-central area, with a center of radiation extending from Wazīristān to Bannū in a northeasterly direction, with Afr, Bg, and Moh the most affected. Afr is by far the best documented; the following five changes are characteristic.

	Psht	Afr
36.13.7.1. <i>á > Afr á</i>		
	<i>las</i> ‘ten’	<i>lās</i>
	<i>yaǵ</i> ‘noise’	<i>yāg</i>

36.13.7.2. $\acute{a} > \acute{\bar{a}}$, also [ɒ:, ɔ:]

<i>plār</i> ‘father’	<i>plār, plōr</i>
<i>pāṇa</i> ‘leaf’	<i>pāṇ(i)ye</i>
<i>bānde</i> ‘upon’	<i>bānde</i>

36.13.7.3. $\acute{o} > \acute{\bar{o}}, \acute{e}$

<i>mor</i> ‘mother’	<i>mōr, mēr</i>
---------------------	-----------------

also $\bar{o} > \bar{e}$ in

<i>bōḡ</i> ‘collected’	<i>bēḡ</i>
<i>špaḡ</i> ‘six’	<i>špēḡ</i>

and sometimes $\bar{o} > u$

<i>xōḡ</i> ‘sweet’	<i>xug</i>
--------------------	------------

36.13.7.4. $\bar{u} > \bar{i}$

<i>ūdā</i> ‘asleep’	<i>wīdā</i>
---------------------	-------------

36.13.7.5. $\acute{a} > \acute{e}$

<i>wraj</i> ‘day’	<i>wrez</i>
<i>wāḡay</i> ‘ear of corn’	<i>wēgay</i>

Note also that quite commonly $o- > wo-$, e.g. Psht (*h*)*óya* ‘egg’ > Afr *wóya*.

36.13.8. *Yūsufzai*

Yus partakes of many of these characteristics as well; note also Yus *nāmā* ‘name’, < Psht *nāmā* (f.), cf. Psht *num* (m.).

36.14. Diphthongs

The clearest way to illustrate the diphthongs of Pashto is to begin with those of the southwest group, where the largest number is to be found, five short and two long, as follows:

short *ay oy uy aw*, and *-əy* which occurs only finally

long *āy āw*

Stress plays a basic role: correspondences in other dialects depend on it. The long diphthongs are always stressed, whilst the short diphthongs may or may not be stressed. Phonetically we have, respectively:

short [aɪ ɔɪ uɪ aʊ], and [əɪ] of 2 morae

long [a:i a:u] of 3 morae

All are stressed on the first member.

36.14.1. *Southwest Group*

Unstressed *əy* = *ay*. In stressed position, however, these two are phonemically and morphologically distinct: in final position nouns ending in *-āy* are

masculines, inflected differently from masculines in unstressed *-ay*; whereas nouns ending in stressed *-ǝy* are feminines, inflected like feminines in stressed *-ǝ̃*.

In Kand *-oy* can be stressed or unstressed. Examples of stressed *-óy* are *toy* ‘overflowing’, *joy* ‘son’. Kand *ow* is probably not distinct from *aw*.

36.14.2. Southeast Group

I can offer only miscellaneous examples. The following have been recorded:

-ǝy > $\left\{ \begin{array}{l} \text{Tar } -(i)y\acute{e} \text{ e.g. } waliy\acute{e} \text{ f. ‘root’} < \text{Psht } wul\acute{o}y \\ \text{Bn } -íy\acute{e} \text{ e.g. } maríy\acute{e} \text{ f. ‘mouthful’} < \text{Psht } nwar\acute{o}y \end{array} \right.$

whereas Waz *pōnyé* ‘leaf’ < **pāṇǝy*, cf. Kand *pāṇa*, with shifted stress.

In Kak, stressed *-ǝy* often becomes *-ǝ̃*, unstressed *´ǝy* = *´ay*, as expected (but *-āy* > Kak *-áy* and *-ǝy*; *-oy* > Kak *-ōe*, e.g. *zōe* ‘son’).

In Waz,

-ǝy > *-áy*: *wray* < Psht *orǝy* ‘gums’; but *-áy* remains, cf. Waz *γīn-dáy* ‘double sack’; *špežmáy* ‘moonlight’; *līgáy* ‘smoke’; unstressed *´ay* remains in *myásay* ‘mosquito’ < Psht *mǎšay*; *-úy* > *-óyǝ* in *nyóyǝ* ‘maternal uncle’ < Psht *nuy*; *-óy* > *-ōi*, *-oy*, *-ōi*; e.g. *tōi*, *tōi* ‘overflowing’ < Psht *toy*; *wǝrbōy* ‘door frame’ < *wǝrbōy*. Note also *-óy* > *-áy* in *zyay* ‘son’ (with palatalised *z*) but this word could be a borrowing from Khost or Paktya dialects.

In Wan morphology, uniquely,

Psht stressed *-áy* masculines correspond to *-á* masculines: *largá* ‘firewood’: Psht *largáy*; *serlá* ‘yearling kid’ Psht *serláy*, etc.; Psht stressed *-ǝy* feminines correspond to *-ǝ̃* feminines: *spožmǝ̃* ‘moonlight’ < *spožmǝy*; Psht unstressed *´ay* masculines correspond to *´a* masculines: *mǝ́na* ‘autumn’: Psht *mǝ́nay*; but unstressed *´ay* feminines (= unstressed *´ǝy*) correspond to *´ǝ̃* feminines: *marǝ́lǝ̃* ‘female companion’: Psht *malǝ́re*.

36.14.3. Northeast and Northwest Groups

-áy > *-ǝ̃*, but often in Afr *-áy*, Moh *-áy* [æɪ], e.g. Moh, Bg *mušáy* ‘fish’ < Psht *mašáy*;

´ay > *-ǝ̃*, and in Afr also *´ay*; but cf. Afr *rǝ́we* < *rǝ́way* ‘a demon’;

-ǝy > *-ǝ̃*, also *-(i)yé*: cf. Afr *pōṇ(i)yé* ‘leaf’ < **pāṇǝy*; cf. Waz,

§ 36.13.5.1.1; Afr *wál(i)ye* < **waliyé* < **wulǎy*, cf. Tar

§ 36.14.2.

-ǎy > -ɛɪ, -ɛɛ, -ǎɪ, -ǎɛ in Hazara;

-uy > -uī, cf. Yus *nuī* ‘maternal uncle’, < *nuy*;

-ǎy > -way in Afr *zway* ‘son’. The long diphthongs -āy > -āɛ and

-āw > -āo in most northeast dialects.

36.15. Vowels and diphthongs in the Pashto script

The Pashto literary language is neither precisely Kand nor Pesh, but rather a mixture of the two.

In contrast to the care taken in writing consonants, vowels and diphthongs in what is basically a consonantal script are much less successfully represented. Dialect writing is extremely unusual; in both Peshawar and Kabul a standard language is the rule in writing, with the short vowels *a i u* ǎ; long vowels *ā ī ū ē ō*.

36.15.1. Short vowels

The system used is derived from Persian, where the short vowels *a, i, u* can be written as “diacritics” above or below the consonant which precedes them, but, as in Persian, they are not usually written at all except in dictionaries. Very occasionally an attempt is made to distinguish ǎ “zwarakay” by the use of ʾ hamza over the preceding consonant. A final short vowel is indicated by -h ʾ, and final -ǎ is sometimes indicated by ʾ hamza over -h (ʾ).

In Afghanistan, the short vowels were formerly sometimes indicated by a plene writing, in which *a* was written *alif*, *i* was written *ya*, and *u* was written *waw*, but this practice, which was not used in Peshawar, has for the most part been abandoned.

36.15.2. Long vowels

Long *ī* (i) and long *ē* (e) have both been written until fairly recently *ی* or *ي*. They were only distinguished by the use of a preceding *kesra*, if at all: a preceding *kesra* meant *ī*, no *kesra* meant *ē*; but practice was never consistent. Recently in Afghanistan the official script adopted since 1958 has had *ی* with two vertical dots as *ē* (e), both internally and in final position. In Peshawar -*ē*- is not distinguished internally from -*ī*- (except as indicated above), but for final -*ē* the Urdu *ے* is used. Long *ō* (o) and *ū* (u) are not distinguished at all, *waw* being used for both. Sometimes a preceding *zamma* is used to indicate *ū*, but this is unusual.

36.15.3. *Diphthongs*

Initial *ay-* is written ای, medially یَ, and finally یِ, with or without *fatha*; in Peshawar final *-ay* is usually written ے.

Final *-əy* has been written in Kabul since the 1958 reform as یر, called “*ya-e fail*”; the older writing was ئی, preserved in Peshawar often as یی.

oy and *uy* are written initially as و ی without distinction, and finally as و ی; in Peshawar final *-uy* is often distinguished as ویی.

Initial *aw-* is written او, and medially and finally و with or without a preceding *fatha*.

Initial *āy-* is written آي, and medially and finally ای.

Initial *āw-* is written او, and medially and finally او.

No account is taken in writing of dialectal variations in diphthong pronunciation.

36.16. Various phonological phenomena

The following is a list of certain typical processes in Pashto. Where possible the dialect is given; for many examples, however, sources are not dialect-specific, and when the forms cited seem credible it seemed a pity to exclude them on that ground alone. But what follows is necessarily somewhat haphazard and certainly incomplete. The main phenomena to be illustrated are palatalization (with remarks and forms of certain relationship-words), metathesis, and dissimilation, striking processes in Pashto for which there is quite a lot of data, even if too often the provenience of the form is unclear.

36.16.1. *Palatalization* (see also §§ 36.11.1, 36.11.2)

All Pashto dialects are strongly subject to palatalization pressures, in vowels or consonants or both, and it seems likely that in some dialects (especially Kak and Wan, on present evidence) an apparent lack of palatalization is in reality a de-palatalization of earlier palatalized forms. The whole subject seems very complex, and it is in many cases unclear whether the palatalization process has occurred already in proto-Pashto, or has supervened only secondarily in the later history of the language. In what follows, “Psht” means standard forms.

Psht		Waz	Kak
<i>oǵá</i>	‘shoulder’	(y)ežá	<i>wažá</i>
<i>úǵa</i>	‘garlic’	<i>wíža</i>	<i>wáža</i>
<i>mášay</i>	‘mosquito’	<i>myášay</i>	<i>máša</i>
<i>óǵka</i>	‘teardrop’	<i>yéška</i>	<i>wóška</i>

<i>šta</i>	‘is’	<i>sta</i>	<i>sta</i>
<i>toğ-</i>	‘to shave’	<i>tež-</i>	<i>tož-</i>
<i>wáğay</i>	‘ear of corn’	<i>wěžay</i>	<i>wážay</i>

Wan, like Kak, very often shows unpalatalized (or depalatalized) forms; many have been listed in Elfenbein 1967: 568ff., 1984: 57, and need not be repeated here. Further examples from other dialects are:

Psht		
<i>ywağ</i>	‘ear’	Northeast (several dialects) <i>yweg</i>
<i>rwaǰ</i>	‘day’	Afr <i>wrez/j</i> ; Bn (w) <i>rez</i> ; Wan <i>wrez</i> ; Waz (w) <i>rəz</i>
<i>špağ</i>	‘six’	Waz <i>špež</i> ; Afr, Bg <i>špeg</i> ; but Wan <i>špōž</i>
<i>žāma</i>	‘jaw’	Waz, Afr, Gh, Bg <i>zāma</i> ; Yus, Pesh <i>ǰāma</i>
<i>zəmay</i>	‘winter’	Northeast dialects <i>ǰəmay</i> , <i>žəmay</i>

It is certain that the forms with *ž* and *ǰ* show a secondary palatalization before an old **i*.

<i>žəba</i>	‘tongue’	Northeast <i>ǰəba</i> ; Southeast, Wan <i>zəba</i>
-------------	----------	--

36.16.1.1. Kin terms

The old inherited words for family relationships show quite a remarkable variety of forms, often within the same dialect.

Psht		
<i>plār</i>	‘father’	pl. <i>plā/arúna</i> in most dialects; <i>plā/arān</i> has also been recorded; Kak children's word sg. <i>p(l)yār</i> ; Wan <i>pyār</i>
<i>mor</i>	‘mother’	pls. <i>máynde</i> , <i>mánde</i> ; Pesh <i>ménday</i> ; Afr <i>máinde</i> ; Kak, NE Gh <i>myénde</i> ; Kak, Moh <i>myánde</i> , Waz <i>máńdyē</i> . The basis of the plural forms must have been, according to Morgenstierne, * <i>māntayah</i> , a possible *Iranian form, which developed into a palatalized <i>m'an'd'ē</i> , before splitting into the various dialect forms.
<i>lur</i>	‘daughter’	plurals <i>lúṇa/e</i> , <i>lúre</i> , <i>lur(g)āne</i> ; Pesh <i>lṇa</i> ; Wan, Kak <i>lúnī</i>
<i>wror</i>	‘brother’	pl. in most dialects <i>wruṇá</i> , as in Kand, Pesh
<i>xwar</i>	‘sister’	Northeast Pesh <i>xwar</i> ; Southeast, Southwest <i>xor</i> ; plurals <i>xwáinde</i> , <i>xwá/and(y)e</i> , <i>xwyánde</i> ; Pesh <i>xwénday</i> ; Wan <i>xūrī</i> , <i>x(w)yénde</i>

<i>zoy, zōe</i>	‘son’	Pesh <i>zóway</i> , Waz <i>zyay</i> , Afr <i>zwa/āy</i> , pl. <i>zoe(y)ǎn</i> ; Pesh pl. <i>zǎman</i>
<i>tandór</i>	‘father’s brother’s wife’	<i>tindǎra, tandyǎra, t(ə)ri/endǎra</i> (all sg.)
<i>tror</i>	‘father’s/mother’s sister’	pl. <i>tránde</i> ; Southwest also <i>trándī, tróre</i> ; Wan <i>trándī</i> ; Central dialects, Waz <i>tránde, tráinde, tryánde, trándye</i> ; Northeast <i>tror(yǎn)e, trénde, trorǎnī, tronǎne</i>
<i>trə</i>	‘father’s brother’	plurals <i>trúna, trúṇa</i>

36.16.2. *Miscellaneous*

<i>or</i>	‘fire’	Northeast, Afr <i>wor</i> ; Wan <i>áwəɾ, owɾ</i> ; unspecified <i>yor</i>
<i>kxe</i>	‘in’, postpos.	most dialects <i>ke</i> ; but Bn <i>še</i> ; Wan <i>če</i>

36.16.3. *Dissimilation*

<i>(n)jən</i>	‘young girl’	<i>njəl(əy)</i>
<i>lǎrya</i>	‘delay’	<i>nǎrya</i>
<i>mlāst</i>	‘prone’	<i>lmōst</i> ; Northeast, Waz <i>nmōst</i>
<i>ngor</i>	‘daughter-in-law’	<i>mğor</i>
<i>nmānj-</i>	‘to honor’	<i>nmāz-, lmānj-</i>
<i>nmunj</i>	‘prayer’	<i>lmunj, munj</i> ; Waz <i>lminz</i>
<i>nwar</i>	‘sun’	<i>nmar, lmar, mar</i> ; Northeast, Southwest <i>nmer, lmer, myer</i> ; Wan <i>mīr</i>

36.16.4. *Metathesis*

<i>maɾwánd</i>	‘wrist’	<i>wəɾmánd</i>
<i>g̃mənɟ</i>	‘comb’	<i>wʒanj, mğanj</i> ; Northeast <i>mangáz</i> ; Wan <i>w(u)žənz</i>
<i>pa(r)xél</i>	‘causing an invalid to sweat’	<i>parxyǎl, paxyǎl, payxál</i>
<i>p(a)sarláy</i>	‘springtime’	<i>parsaláy, sparláy, sarláy</i> ; Wan <i>swarlá</i>
<i>p̃xa</i>	‘foot’	Northeast <i>xpa</i> ; Wan <i>špa</i>
<i>rwaj</i>	‘day’	Bn, Wan <i>(w)rez</i> ; Pesh <i>wraj</i> ; Waz <i>(w)rəz</i>
<i>aɣzáy</i>	‘thorn’	<i>azyáy</i>

byā/arg ‘doubled, folded’

ybarg, bray

Kand, Waz, Bn *yağ*,

Gh *yaž* ‘sound, noise’ Afr *yāg*, Khost *žāy*, Wan *žay*

lār ‘road’ < **rāl* < *Ir *rāθī*-

lwan- ‘to scatter’ Pesh *nwal-*

mayzáy ‘border, hem’ Kak *mazyáy*

māzyé ‘brain’ *māyzé*; Waz *moyzé*

mzaráy ‘tiger’ *zmaráy*; Kak *mjaráy*; Wan *mzará*

riná ‘awl’ *nirá*

spoğmáý ‘moon(light)’ *spom/nğóý*

sxay ‘young calf’ Afr, Waz *x(u)sáy*

sxar ‘father-in-law’ Waz, Kak, Tar *xwsár*

trix m., *tərxá* f.

‘bitter’ Pesh *traxá* f.

txarg ‘armpit’ *trax*

ǵkul ‘kiss’ *kǵul*

ǵax ‘hard, stiff’ *xaǵ*

zmǵka ‘earth’ Waz, Wan *mzǵka*

This is only a small sample.

36.17. Loanword phonology

36.17.1. Early loanwords from NP/Ar

š > *ǵ*, *ž* > *ğ*, *č* > *c*, *ǵ* > *j*; the following are early loanwords:

Psht

NP

laǵkár ‘army’ *laškar*

koǵíǵ ‘attempt’ *kōšiš*

peǵ ‘happening’ *pēš* (*āmadan*)

wreǵám ‘silk’ *abrēšam*

doǵáx ‘hell’ **dōžax* < *dōšax*; but Parthian *dōžax*

noǵ ‘pure’ **nož*, < *nōš* ‘sweet’

cáwda ‘cave’ *čafta*

carx ‘circle’ *čarx*

car- ‘to graze’ *čarīdan*

jawáb ‘answer’ *ǵawāb*

jigár ‘liver’ *ǵigar*

duǵmón ‘enemy’ *dušman*

36.17.2. Palatals in late loanwords

In late loanwords from NP/Ar, *š*, *ž*, *č*, *ž* remain.

<i>šmer-</i>	‘to reckon’	< NP <i>šumarīdan</i>
<i>pišī</i>	‘cat’	NP dialects <i>pōšak</i> , Balochi <i>piššī</i>
<i>šalōy</i>	‘rice’	NP dialects <i>šālī</i>
<i>bačāy</i>	‘child’	NP dialects <i>bača</i>
<i>məč</i>	‘fly’	< IA
<i>čāp</i>	‘printing’	< NP (< IA)
<i>čāwṇōy</i>	‘military camp’	< IA
<i>jism</i>	‘body’	NP/Ar
<i>žārú</i>	‘broom’	NP
<i>kunžī</i>	‘key’	Si/Lhd

36.17.3. Stops in late loanwords

In late loanwords the stops *b*, *d*, *g* remain; e.g. from IA *bal-* ‘to call’; *duk* ‘affliction’; *gor-* ‘to see’; *gaṛāy* ‘jug’; *bučšray* ‘donkey foal’; *dobī* ‘washer-man’, etc.

Loanwords with aspirates in IA lose their aspiration in loanwords in Pashto; but IA *t*, *ḍ*, *r*, *ṇ* are kept.

Psht		
<i>koṭ</i>	‘fort’	< Lhd
<i>ṭokrōy</i>	‘blanket’	< Si
<i>ḍawl</i>	‘manner’	< Lhd
<i>gəḍ</i>	‘sheep’	< Lhd
<i>kaṛōy</i>	‘ring’	< Si/Lhd
<i>gidār</i>	‘fox’	< Pnj

(but cf. *-ḍ-* < IA *-ṛ-* in *baḍāy* ‘rich man’)

<i>gaṇ-</i>	‘to count’	< Pnj
<i>gaṇōy</i>	‘churn’	< Lhd

36.17.4. Miscellaneous

NP *-rd* > *-ṛ* in the early loanword *gogəṛ* ‘match’, but > *-ṛd* in the late loanwords *daṛd* ‘pain’, *paṛdā* ‘curtain’.

f from whatever source generally > *p*, e.g. *plānāy* ‘someone’ < NP/Ar *fulān*; *prānsé* ‘the French language’; *píptī-píptī* ‘fifty-fifty’, etc. But cf. *fulél* ‘hair oil’, a hyper-Arabism, < Si *phulelu*.

Bibliography

- Barth, F. 1964. "Ethnic Processes on the Pathan–Baloch Boundary." In *Indo-Iranica: Mélanges ... Morgenstierne*. Wiesbaden: Harrassowitz.
- Darmesteter, James. 1888–90. *Chants populaires des Afghans*. Paris.
- Elfenbein, J. 1967. "Laṇḍa, zor wəla." *Archiv Orientální* 35: 563–606.
- . 1984. "The Wanetsi Connexion." *Journal of the Royal Asiatic Society* 54–76, 229–41.
- Gershevitch, Ilya. 1959. *The Avestan Hymn to Mithra*. Cambridge: Cambridge University Press.
- Grierson, G. A. 1921. *Linguistic Survey of India*, vol. 10, *Eranian Family* (repr. Delhi, 1968) (bibliography on pp. 14–16).
- Kieffer, Charles. 1974. "L'établissement des cartes phonétiques, premiers résultats, L'atlas linguistique des parlers iraniens." *Atlas de l'Afghanistan* (Universität Bern, Arbeitspapiere 13).
- . 1983. "Afghanistan. v. Languages." *Encyclopaedia Iranica*, vol. 1, pp. 501–16.
- Lorimer, J. G. 1902. *Grammar and Vocabulary of Waziri Pashto*. Calcutta.
- Loy, Lucia. 1987. *Il Tesoro nascosto degli Afghani*. Bologna: Il Cavaliere Azzurro.
- MacKenzie, D. N. 1959. "A standard Pashto." *Bulletin of the School of Oriental and African Studies* 22: 231–35.
- . 1964. "The Xayr ul-bayān." In *Indo-Iranica: Mélanges ... Morgenstierne*, pp. 134–40. Wiesbaden: .
- . 1965. *Poems from the Divan of Khushal Khan Khattak*. London.
- . 1987. "Pashto." in *The World's Major Languages*, ed. Bernard Comrie, pp. 547–65. London: Croom Helm.
- Morgenstierne, Georg. 1927. *Etymological Vocabulary of Pashto*. Oslo.
- . 1930. "The Wanetsi dialect of Pashto." *Norsk Tidsskrift for Sprogvidenskap* 4: 156–75 (repr. in *Irano-Dardica* [Wiesbaden, 1973], pp. 187–206).
- . 1931. "The Story of an Afridi Sepoy." *Studia Indo-Iranica, Ehrengabe für W. Geiger*, Leipzig (repr. in *Irano-Dardica*, pp. 175–86).
- . 1932. *Report on a Linguistic Mission to Northwestern India*. Oslo.
- . 1940. "'Pashto', 'Pathan', and the treatment of *r* + Sibilant in Pashto." *Acta Orientalia* 138–44.
- . 1942. "Archaisms and Innovations in Pashto Morphology." *Norsk Tidsskrift for Sprogvidenskap* 12: 88–114.

- . 1948. "The development of *r* + Sibilant in some Eastern Iranian Languages." *Transactions of the Philological Society* 70–80.
- . 1950. "Additional notes on 'The Development of *r* + Sibilant'" *Transactions of the Philological Society* 207.
- . 1960. "Afghān." *Encyclopaedia of Islam*, 2nd ed., vol. 1, pp. 216–21.
- . 1973. "Additional Notes on Wanetsi." *Irano-Dardica*, pp. 207–23.
- . 1983a. "Afghanistan. vi. Pašto." *Encyclopaedia Iranica*, vol. 1, pp. 516–22.
- . 1983b. "Bemerkungen zum Wort-Akzent in den Gathas und im Paschto." *Münchener Studien zur Sprachwissenschaft* 42: 167–75.
- Penzl, Herbert. 1955. *A Grammar of Pashto*. Washington, D.C.: American Council of Learned Societies.
- Skalmowski, W. 1975. "Two Stories in Afridi Dialect." *Acta Iranica* 5: 243–49.
- Skjærvø, Prods O. 1989. "Pashto." In *Compendium Linguarum Iranicarum*, ed. Rüdiger Schmitt, pp. 384–410 § 4.2.1. Wiesbaden: Reichert (with a good bibliography pp. 409–10).
- Vercellin, Giorgio. 1983. "AWQ." *Quaderni del Seminario di Iranistica, Uralo-Altaistica, e Caucasiologia dell'Università degli Studi di Venezia*, No. 21, Parte sesta, pp. 1–25.

Balochi Phonology

Josef Elfenbein

Johannes Gutenberg–Universität Mainz

37.1. Introduction

Balochi (Balōčī, conventionally Baluchi) is an Iranian language of the North-West group, with closest linguistic affinities to e.g. Kurdish, Tālīšī, and Tātī, and whose closest Middle Iranian ancestor was Parthian. But no known Middle Iranian language is a direct ancestor of Bal(ochi), whose present situation in the extreme southeast of the Iranian-speaking area is due to migration since late Sassanian times (from ca. 600 A.D.) from an earlier habitat in the Caspian area of north-central Iran. (For a fuller discussion of these matters see Spooner 1989; this article also includes a full bibliography of previous writings on the subject. There is also a brief summary in *CLI* § 4.1.2.6.1.2.)

Balochi is spoken principally by Baloch tribesmen, as well as perhaps two-thirds of Brahui tribesmen (who speak it as a mother tongue); but by no means all of those describing themselves as Baloch speak Balochi even as a subsidiary language—such Baloch live mainly in Sind and in the Gulf States. The main Balochi-speaking area is a huge one, however, extending as it does from Marw in Turkmenistan southwards in pockets through Iranian Khorāsān, and then throughout the province of Balūčestan-va-Sīstān in Iran to the Gulf of Oman; in Afghanistan Balochi-speaking areas extend from Sīstān in the west along the Hilmand Basin eastwards to Šōrāwak at the Pakistan frontier; in Pakistan throughout the province of Baluchistan, including a salient to the north-east as far as Dera Ismail Khan in the North-west Frontier Province; large numbers of Balochi speakers also live in Karachi and in other large cities in Sind (see *CLI* § 4.1.2.6.1.3 for more detailed information and estimates of the number of speakers: to the information

AUTHOR'S NOTE: Abbreviations: Af, Afghan; Ar, Arabic; *CLI*, *Corpus Linguarum Iranicarum*; Co, Coastal; EHB, Eastern Hill Balochi; g., genitive; IA, Indo-Aryan; intr., intransitive; Ke, Kēčī; La, Lāšārī; pl., plural; p.p., past participle; pres., present; NP, New Persian; Ra, Raxšānī; Sa, Sarāwānī; sg., singular.

provided there about the number of speakers in Turkmenistan should be added information kindly supplied to me by Prof. A. L. Gryunberg in 1990: the number of Balochi speakers is of the order of 25–30,000). The total number of Balochi speakers must remain, in the absence of any comprehensive reliable census data, somewhat uncertain, but I feel that the estimates given in the above-cited articles of ca. 3,600,000 can stand as an upper limit; it may be perhaps 10% on the high side.

37.2. Publishing in Balochi

Balochi has never been an official language anywhere—until recently in Afghanistan after 1979, when the Tarakī Government declared as “national” languages Balochi, Kati, Pashai, Uzbek, and Turkmani (the “official” languages are Persian and Pashto). The main agency responsible for Balochi publications in Kabul is the *Zuwān wa adabiyātē institut*, which has published books and a newspaper (*Sōb* ‘Victory’). The main centre of Balochi cultural and literary activity lies however outside Balochi-speaking territory in Quetta, Pakistan, where publishing in the language (books and magazines) began in the 1930s, and has been especially active since 1950. The chief publishing agency in Balochi is the Baluchi Academy, with an impressive list of publications in Balochi to its credit, mainly original work. Karachi has also seen the establishment (and demise) of several Academies for the same purpose: the most important now is the Sayid Hašmī Akadēmī (for a fuller account of these matters see *CLI* § 4.1.2.6.2, Elfenbein 1989b: 637; Jahani 1989 has a useful summary in chap. 5).

37.3. Balochi dialects

Many Balochi speakers speak a second or third language as well—Brahui, Sindhi, Sindhi-Siraiki, Persian, Urdu, Pashto—depending on their habitat and social connections. Since the phonology of the various dialects differs, a brief description of them is necessary. They are six in number, and with one exception they are all mutually intelligible. An account of their main distinguishing characteristics has been given in *CLI* § 4.1.2.6.6, and in Elfenbein 1989b: 636–37; the most detailed account is to be found in Elfenbein 1990, vol. 2, pp. vii–xviii. The following summarizes the phonological aspects of interest here; the transcription used is the most common; for details see below, § 37.4.

37.3.1. *Raxšānī*

Old *ū* is kept unchanged.

Old past participles in **-xt-* show either *-ht-* in Kalātī/Panḡgūrī/Sarhaddī (in Afghanistan, Sīstān, Marw) or *-t-* in Afghan Ra, with compensatory vowel lengthening.

Nasalization of vowels is avoided, and postvocalic *-n* is stable.

h is exceedingly unstable in all positions: it is often preserved in southern Ra, whereas it is dropped in all positions in AfRa.

Initial *gwa-* usually appears as *gu-*.

An epenthesis between two back vowels in close juncture appears usually in AfRa as *-w-*, but elsewhere in Ra as either *-w-* or *-y-*.

37.3.2. *Kēcī*

Old *ū* is kept unchanged.

Old past participles in **-xt-* show a metathesis (with **x > k*) to *-tk-*.

Nasalization of vowels is avoided, and postvocalic *-n* is stable.

h is stable.

Initial *gwa-* is often kept, and as often reduces to *gu-*.

An epenthesis between two back vowels in close juncture usually appears as *-h-*.

37.3.3. *Coastal dialect*

A prestige dialect for epic poetry and other literature, often imitated by others.

Old *ū* is fronted to *ī*, except in very recent loanwords.

Old past participles in **-xt-* show a metathesis (with **x > k*) to *-tk-*.

Vowels preceding an *n* are usually nasalized.

h is stable, and there is a tendency to prefix it unetymologically to some words with an initial vowel.

Initial *gwa-* is usually kept unchanged.

An epenthesis between two back vowels in close juncture is with *-h-*.

37.3.4. *Sarāwānī*

A prestige dialect in Iran.

Old *ū* is usually fronted to *ī* in stressed syllables.

There is a mild drawl affecting stressed *ō* and *ē*, so that most speakers pronounce them [u̯o], [i̯e] respectively; some speakers tend to adopt [a:] for [a:] in the Persian manner.

Old past participles in *-*xt-* are mainly retained as -*xt-*, but there is a tendency with some speakers to develop this to -*ht-*.

Nasalization of vowels preceding *n* is sporadic.

In suffixed morphemes with final -*n*, the preceding vowel is often nasalized and the *n* dropped.

h is stable, and there is a tendency to prefix unetymological *h* to words with an initial vowel.

Initial *gwa-* is usually kept.

The preferred epenthesis between two back vowels in close juncture is with -*y-*.

37.3.5. *Lāšārī*

Old *ū* is often kept unchanged; but some speakers (probably influenced by Sa, the prestige dialect in Iran) tend to front it to *ũ*; and the *ō* of other dialects is often heard as *ō̃* [ø:]. *ū* and *ō*, however, often appear in recent loanwords from Persian, and when stressed, /*ū*/ is strongly drawled to *ūō* [uó:] and /*ē*/ to *ēē* [ié:].

Past participles in old *-*xt-* show (with **x* > *k*) a metathesis to -*tk-*.

Vowels preceding *n* are often nasalized, and in particular -*ān*, -*ā̃* > -*ōn* or -*ō̃*; the nasalising *n* is usually dropped.

h is stable, and unetymological *h* is often heard.

Initial *gwa-* is preferred to *gu-*.

Epenthesis between two back vowels in close juncture is usually made with -*h-*.

37.3.6. *Eastern Hill Balochi*

This dialect, with many sub-varieties, is the most complex of all in its phonological structure, as a consequence of which (as well as its morphology and vocabulary) it is with difficulty intelligible to speakers of the other dialects.

Prevocalic voiceless stops and affricates (esp. in initial position) are audibly more aspirated than in other dialects, but this aspiration is never phonemic. The voiced aspirates *bh gh dh jh* (only in loanwords from Indo-Aryan) do not contrast with *b g d ĵ*. The postvocalic stops and affricates *k p t g b d č ĵ* are opened to *x f θ γ β δ s z* respectively, except in pre-consonantal position. For the (occasional) aspiration of *w*, see below.

The voiceless fricatives *š f θ x* can be voiced in postvocalic position.

All final postvocalic consonants can be dropped.

Final double consonants are usually simplified.

Pre-consonantal *r* usually becomes *ɾ*.

Old *ū* is fronted to *ī* except in very recent loanwords.

The ending *-ōn* usually becomes *-ñ*.

Old past participles in **-xt-* are usually kept, though side by side with *-tk-* forms, depending on the speaker.

Vowels preceding *n* are strongly nasalized, with a consequent dropping of *n* except when a vowel follows.

h is usually stable.

gwa- is kept.

Epenthesis is with *-h-* between two back vowels.

37.4. General phonology

In the following, all dialects are treated together.

37.4.1. Vowels

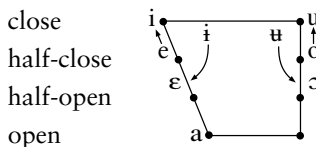
Balochi has at base 8 segmental vowel phonemes in an asymmetrical pattern:

short	<i>i</i>	<i>u</i>	<i>a</i>		
long	<i>ī</i>	<i>ū</i>	<i>ā</i>	<i>ē</i>	<i>ō</i>

with non-phonemic variations, a system almost identical to that of Western Middle Iranian (for example, see *CLI*, § 3.1.1.4).

For altered vowels in Sa, La see §§ 37.3.4, 37.3.5.

The vowel quadrilateral is:



Phonologically, *e* and *o* are always long, and it is unnecessary to write them with macrons in phonemic writing; but to avoid confusion, macrons are utilized here.

Note that the long vowels *ī* [i:] and *ū* [u:] admit of no variation, and are in pronunciation entirely stable in all dialects. Those dialects in which *ī* has developed from old *ū* do not differentiate the developed *ī* from an original *ī*. Long *ā* tends to [a:] in some styles of Sa and La, but the variation is non-phonemic.

$\bar{i}$ [i:] is a very close, very long front vowel (2–3 morae) in all positions; $\bar{u}$ [u:] is a very long, very rounded back vowel (2–3 morae) in all positions; $\bar{a}$ [a:] is an open centralized very long unrounded vowel (2–3 morae) in all positions. In the dialects Sa, La it is freely rounded to [ɑ:] in stressed position.

i [i] is a lowered close unrounded front short vowel (1 mora); in stressed syllables it tends to [ɛ]. It is somewhat unstable.

u [ʊ] is a lowered close rounded short vowel (1 mora). In stressed syllables and before r it tends to [ɔ]. Somewhat unstable.

a [a] is an open centralized short vowel (1 mora). Somewhat unstable, it tends in unstressed syllables to [ɐ], especially in EHB.

The contrasts $i : \bar{i}$ and $u : \bar{u}$ are thus both qualitative and quantitative, whilst the contrast $a : \bar{a}$ is mainly only quantitative (except EHB), and some speakers of Sa and La contrast $a : \bar{a}$ as [a] : [ɑ:]. The contrast of $a : \bar{a}$ is important:

<i>man</i> : <i>mān</i>	‘I’ : ‘in’
<i>kad</i> : <i>kād</i>	‘manure’ : ‘bite’
<i>pad</i> : <i>pād</i>	‘trace’ : ‘foot’
<i>dar</i> : <i>dār</i>	‘outside’ : ‘wood’
<i>ač</i> : <i>āč</i>	‘from’ : ‘fire’
<i>čar-</i> : <i>čār-</i>	‘to graze’ : ‘to look’

Other minimal pairs are:

$i : \bar{i}$	<i>bin</i> : <i>bīn</i>	‘larynx’ : ‘a kind of flute’
	<i>bīr</i> : <i>bīr</i>	‘on, upon’ : ‘expert, champion’
	<i>šīr</i> : <i>šīr</i>	‘lentil’ : ‘milk’
$u : \bar{u}$	<i>jur</i> : <i>jūr</i>	‘cloud bank’ : ‘lower intestine’
	<i>gul</i> : <i>gūl</i>	‘rose’ : ‘demon’

The only case of a consistent allophonic pronunciation of i is in the morpheme for the 2 pl. verbal ending *-it*, where especially in Ra it is consistently pronounced [-et].

$\bar{e}$ [e:] a tense, mid-front unrounded long vowel (3 morae), somewhat closer than cardinal e , and kept distinct from $\bar{i}$ in all dialects.

$\bar{o}$ [o:] a mid-back rounded long vowel (3 morae), somewhat closer than cardinal [o], and kept distinct from $\bar{u}$ in all dialects.

Historically both $\bar{e}$ and $\bar{o}$ are long vowels. When it is important to distinguish allophonic short e or o , one can write $\check{e}$, $\check{o}$.

Minimal pairs:

<i>bīr</i> : <i>bēr</i>	‘expert, champion’ : ‘revenge’
<i>lōnd</i> : <i>lūd</i>	‘oval’ : ‘crawl on all fours’

<i>šīr</i> : <i>šēr</i>	‘milk’ : ‘lion’
<i>ǰūr</i> : <i>ǰōr</i>	‘lower intestine’ : ‘prepared’
<i>sīl</i> : <i>sēl</i>	‘seal’ : ‘rennet’
<i>kū(h)</i> : <i>kō(h)</i>	‘water well’ : ‘mountain’
<i>kūš</i> : <i>kōš</i>	‘spongy marrow bone’ : ‘murder’
<i>gwamīš</i> : <i>gwamēš</i>	‘small plant used as soap’ : ‘buffalo’

However, in monosyllables final *-ī* and *-ē* are free variants:

<i>čī</i> , <i>čē</i>	‘what, thing’
<i>ē</i> , <i>ī</i>	‘this’

as are, apparently, *-ū* and *-ō* in final position:

<i>gwandō/ū</i>	‘baby’
<i>bō/ū</i>	‘smell’
<i>čō/ū(n)</i>	‘how, thus’
<i>tambō/ū</i>	‘drum’
<i>mātō/ū</i>	‘mother-in-law’

37.4.2. Diphthongs

There are phonemically only two diphthongs, best written /ay/, /aw/, where *ay* = [æi] and *aw* = [ɔu]; both have stress on the first member, with 2–3 morae, and no friction. *ay* is subject to very great free variation in all dialects, especially in Ra: we hear North Ra [a:i, a:i, ai:], South Ra [a:ɛ, a:i, æɛ, æi, ɛi]; most noticeably in final position one hears in Ra [æɛ, æi, ɛi, ɛi, ɛi], even monophthongized to [e:]. /ay/ appears most opened [a:i] in North Ra territory—i.e., in Afghanistan, with 3 morae.

It becomes closer in the South, and is mostly a monophthong [e:] in the Co dialect region. Thus, *ay* and *ē* contrast most sharply in Ra (e.g., *baray* ‘thou bearest’ : *barē* ‘once’), whereas in Ke, Co one hears mostly only *barē* for both.

Mutatis mutandis, with the same geographical distribution we have for the diphthong *aw* [ɔu, ɔu, ɔu, o:] going from North to South, though perhaps in a less marked manner. Contrast *aw* : *o* in Ra : other dialects (but to a lesser extent also in Ke and Sa, and not at all in Co or EHB).

Ra : Other dialects

<i>taw</i> : <i>tō</i>	‘thou’
<i>say</i> : <i>sē</i>	‘three’
<i>šayr</i> : <i>šēr</i>	‘poem’
<i>sayl</i> : <i>sēl</i>	‘walk’

<i>hawr</i> : <i>hōr</i>	‘rain’
<i>aw</i> : <i>ō</i>	‘and’
<i>hayr</i> : <i>hēr</i>	‘welfare’
<i>kawš</i> : <i>kōš</i>	‘breeze’
<i>-ay</i> : <i>-ē</i>	g. sg.; in Ra g. sg. : <i>ya-e wahdat</i>
<i>lawn</i> : <i>lōn</i>	‘kind, quality’
<i>dawl</i> : <i>dōl</i>	‘type, manner’

As a counterpart to the contrast of final *-ay* : *-ē* in Ra, I have not found any Ra minimal pairs for final *-ō* : *-aw* (there are a large number of words in final *-ō* : *-aw*, mainly loanwords from Sindhi: e.g., *kulaw* : *kulō* ‘message’—the former in Ra, the latter in the south).

37.4.3. Epenthesis

37.4.3.1. Consonant epenthesis

Balochi has no glottal stops except as occasional “elegances” in loanwords from Arabic or New Persian; and apart from diphthongs (§ 37.4.2), vowels in close juncture (internally before morphological endings, externally in word juncture) are articulated separately. When long or short *ǎ* is the preceding member of a cluster, a consonantal glide develops before a following vowel, /y/, /w/, or /h/ depending on dialect. The most important such junctures with their dialect distributions are:

<i>ǎ</i> + <i>ǎ</i>	<i>ǎwǎ</i> in AfRa <i>ǎyǎ</i> in South Ra, Ke, Sa <i>ǎhǎ</i> in EHB, Co, La, often simply <i>ǎǎ</i>
<i>ǎ</i> + <i>ũ</i>	<i>ǎwũ</i> , almost <i>ǎũ</i> in AfRa <i>ǎyũ</i> in South Ra, Ke, Sa <i>ǎhũ</i> in EHB, Co, La, often simply <i>ǎũ</i>
<i>ǎ</i> + <i>ō</i>	<i>ǎwō</i> , <i>ǎō</i> in AfRa <i>ǎhō</i> , <i>ǎō</i> elsewhere
<i>ǎ</i> + <i>ě</i>	<i>ǎwě</i> in AfRa <i>ǎyě</i> in South Ra, Ke, Sa <i>ǎhě</i> in EHB, Co, La, often simply <i>ǎě</i>
<i>ǎ</i> + <i>ĩ</i>	<i>ǎyĩ</i> in Ke, Co <i>ǎĩ</i> elsewhere

Other vowel conjunctions are articulated separately with stress on the first member: *ē* +, *ĩ* +, *ō* +, *ũ* +, with no epenthesis.

Exception: *ē* + *ē* > *īē* [í:je:], with stress on the first member.

37.4.3.2. *Vowel epenthesis*

Consonant clusters are sometimes opened, the vowel used mostly being *a*. Especially notable is the development of *a* between a word-final stop and a word-initial stop: -K + K- (where K is any stop) is sounded /-K a K-/; e.g., Ra *manī pādānā singān dard kurtant* ‘the stones hurt my feet’ sounds /manī pādānā singān a dard a kurtant/; Sa /singān manī pād a laggit/ ‘the stones struck my foot’; Ke /āyā asp a boḷiya kutant/ ‘he put the horses on (board) the ship’ Co /man buz a zurtant, a kuštant/ ‘I took the goats and killed (them)’. Typical consonant clusters without epenthesis are:

<i>dračk</i>	‘tree’
<i>gwāt</i>	‘wind’
<i>ḡwān</i>	‘young, good’
<i>dištār</i>	‘bride’
<i>dakhāna</i>	‘post office’ (also <i>dakxāna</i>)
<i>sayrjan</i>	‘singer’
<i>ḡupt</i>	‘pair’
<i>rōdrātk</i>	‘sunup, east’
<i>sitk</i>	‘confidence’

Initial consonant clusters involving *s* or *š* are not common, and usually not opened:

<i>spantān</i>	‘wild rue’
<i>spēt</i>	‘white’
<i>srōšk</i>	‘elbow’
<i>srap</i>	‘line, row’
<i>sruš</i>	‘to knead’
<i>skēn</i>	‘goaded’
<i>špārḡ</i>	‘slim’
<i>šrapar</i>	‘moustache’
<i>šlap</i>	‘splash’

37.4.3.3. *Vowel clusters*

<i>druāī</i>	‘health’
<i>tūt</i>	‘he is huge (3 sg.)’
<i>tūt</i> [tu:it]	‘you are huge (2 pl.)’
<i>mullāē</i>	‘a mullah’
<i>bōē</i>	‘a smell’
<i>čāān</i>	‘tea (pl.)’

<i>dōmīēn</i>	‘second’
<i>tālīāy</i>	‘of the plate (g. sg.)’
<i>dunyāay</i>	‘of the world (g. sg.)’
<i>giraō</i>	‘security’
<i>čīān</i>	‘things’
<i>čēān</i>	‘tea (pl.)’

37.4.3.4. #*gwa*- > #*gu*-

The initial cluster *gwa*- becomes sporadically *gu*- in all dialects, especially in Ra and EHB; but unusual in Co, La:

<i>gwamz</i>	‘wasp’; EHB also <i>gumz</i>
<i>gwahār</i>	‘sister’; sporadically <i>guhār</i> everywhere
<i>gwanjisk</i>	‘sparrow’ (prob. a loanword from Persian with ‘false’ <i>gw</i> -); EHB <i>gunjisk</i>
<i>gwar</i>	‘breast’; <i>gur</i> , <i>gor</i> in EHB, Ra
<i>gwaš-</i>	‘speak’; Ra, Ke <i>guš-</i>
<i>gwaz-</i>	‘pass’; Ra <i>guz-</i>
<i>gwahr</i>	‘cold’; EHB also <i>guhr</i>

37.4.3.5. -*r*-

In Ke and Co an intrusive -*r*- appears in

<i>čar ēdā</i>	‘from here’ (for <i>ča ēdā</i>)
<i>čar ē</i>	‘from this’ (for <i>ča ē</i>)

37.4.3.6. *Nasalization*

All vowels, short and long, may be nasalized. Except as noted below, nasalization is a secondary development from vowel + *n* (in many cases from vowel + *z*, e.g. *pōz/pōnz/pōž* ‘heel’); the tendency to nasalise is strongest in the south and east (Co, EHB) and weakest in the north (AfRa). Southern Ra uses mainly -*ī*, -*ō̃*, -*ū* for the endings -*in*, -*ōn*, -*un*. Nasalization has progressed furthest in EHB, where there is the strongest tendency to use *Ũ* for *V + n* in all positions, except when a vowel follows.

Examples: *gōn* ‘with’, EHB *gō*; *šarrēn* ‘fine’, all dialects except Ra *šarrē̃*; in EHB before *z*: *dāz* ‘dust’, other dialects *danz*; *sirōnz* ‘fiddle’, EHB *sirō̃z*; -*int*, 3 sg. pres. ending, EHB -*ī̃*; -*ant*, 3 pl. pres. ending, EHB -*ā̃*; verbal endings -*un*, -*an*, -*ēn*, EHB, Co -*ū̃*, -*ā̃*, -*ē̃*; but in all dialects the numbers 11–16 and 19 are *yāzda*, *dwāzda*, *sēzda*, *čārda*, also *čyārda* (EHB), *čārda* (Ke, Co), *pāzda*, *šāzda*, *nōzda*, also *nōzda* (Ke, Co).

37.4.4. Segmental consonant phonemes

Table 37-1. Balochi Consonants

	Labials	Dentals	Palato-Alveolars	Retroflexes	Velars	Glottals
Stops	p b (bh)	t d (dh)		ʈ ɖ (ɖh)	k g (gh)	ʔ
Affricates			tʃ ɟʃ			
Fricatives	f	(θ) (ð)			x ɣ	h
Spirants		s z	ʃ ʒ			
Nasals	m	n		(ŋ)	(ŋ)	
Lateral		l				
Flaps		r		ɽ		
Semivowels	w	y				

Bracketed sounds are explained below.

p [p] voiceless bilabial stop, slightly aspirated prevocally and finally. Only in EHB strongly aspirated initially as a rule [ph], but with many exceptions, e.g., EHB [phull, niphad], but [pōh, phōh; pa, pha; pōsay, phōsay; prāh, phrāh; phisphairī, pispairī, phispairī]. The aspiration is non-phonemic. *p* is developed from *f* in loanwords in all dialects, except in “elegant” speech.

b [b] voiced bilabial stop, unaspirated. *bh* is kept in loanwords from IA in EHB as a rule, but is not phonemic. [bhōrēnay, sambharta], cf. [bōrēnag, sambarta] in other dialects.

t [t] voiceless post-dental stop, in EHB and other dialects slightly aspirated prevocally and finally. Only in EHB as a rule strongly aspirated in initial position [th]. The aspiration is non-phonemic. In EHB finally, post-vocally, > Ø. Cf. EHB [thāwān, thāwā, thēlank]: other dialects [tāwān, tēlank]. Non-EHB [kapīt, gušant]: EHB [kapī, gušā].

d [d] voiced post-dental stop, unaspirated. Initial *dh* from IA loanwords is kept as a rule only in EHB [dhōl, dhakk], cf. other dialects [dōl, dakk]. *dh* is non-phonemic in EHB.

ʈ [ʈ] voiceless retroflex stop, unaspirated. Occurs in non-EHB only in loanwords < IA. In EHB aspirated initially as a rule: EHB [ṭēsan, ṭhēsan, ṭōr, ṭhōr]: other dialects [ṭēsan, ṭōr]

ḍ [d̥] voiced retroflex stop, unaspirated. Occurs in non-EHB only in loanwords < IA. EHB keeps *ḍh* in IA loanwords as a rule [ḍhōnd, ḍhōndh], cf. [ḍōnd, ḍūnd]. Intervocally it can interchange with *-ṛ-* in Ra (see under *ṛ*).

k [k] voiceless velar stop, slightly aspirated finally. Slightly fronted before front vowels, backed before back vowels. In EHB as a rule aspirated [kh] prevocally: [khanay, khār]; but also [kōh, khōh; kaššay, khassay; kass, khass, kamm, kamm].

g [g] voiced velar stop, unaspirated. Slightly fronted before front vowels, backed before back. EHB keeps as a rule [gh] from IA loanwords in initial position [ghōb, ghaṛō], cf. other dialects [gōb(ī), gaṛō]. *gh* is non-phonemic in EHB.

č [tʃ] voiceless palato-alveolar affricate. It is a single phoneme, since it can be doubled; other affricate groups, *pf*, *ts*, etc., do not occur in initial position; the clusters *čy-*, *čw-* occur initially, and no 3-consonant phoneme groups occur in initial position. Usually aspirated in initial prevocalic position in EHB [čh]. Often kept in EHB in loanwords < IA.

ǰ [dʒ] voiced palato-alveolar affricate, unaspirated. A single phoneme, for the same reasons as under *č* (q.v.). EHB often keeps aspirated *ǰh* in loanwords from IA especially in initial position.

f [f] voiceless labio-dental fricative. In principle only in recent loanwords from Ar/NP, kept in all non-EHB dialects only as an “elegance”; otherwise > *p*. In EHB, regularly < postvocalic *p*.

s [s] voiceless post-dental spirant in all dialects. In some EHB dialects it develops from a postvocalic *θ* (q.v.).

z [z] voiced post-dental spirant in all dialects. In some EHB it develops from a postvocalic *ḍ* (q.v.).

θ [θ] voiceless post-dental fricative. Peculiar to EHB, it develops regularly from a postvocalic *t*, but not always: p.p. in *-(i)ta* develop in EHB into both *-(i)θa*, *-(i)tha* in apparent free variation.

ḍ [ḍ] voiced post-dental fricative. Peculiar to EHB, it develops regularly from a postvocalic *d*, but not always. Note in particular “elegant” EHB *naḍar*, *ḍāt* (not *(ḍaθ!)* with *ḍ* < *z* in Arabic words spelt with *ḍal*.

š [ʃ] voiceless palato-alveolar spirant in all dialects. In EHB, it develops regularly as an allophone of postvocalic *č*.

ž [ʒ] voiced palato-alveolar spirant in all dialects. In EHB it develops regularly from a post-vocalic *ǰ*.

x [x] voiceless velar fricative, mainly in late loanwords from Ar/NP; early loanwords have *h* (cf. late *xudā*, early *hudā*). In EHB it develops regularly from a post-vocalic *k*.

y [ɣ] voiced velar fricative in all dialects, it appears only in loanwords from NP/Ar. In EHB, it develops regularly from post-vocalic *g*.

h [h] voiceless glottal fricative, but with very little friction in all dialects. Very unstable in Ra, it is dropped in all positions in AfRa, and occasionally in other Ra. In Ke, Co, Sa, EHB it is prefixed unetymo-logically in *hasp* ‘horse’, *hask* ‘gazelle’, *harōs* ‘wedding’, *hawr* ‘rain’, and a few other words. For its use in epenthesis, see § 37.4.3; for more details about its stability see § 37.3. *h* tends to be stabler initially before a short vowel: South Ra *hagāzī* ‘musician’, but *āk* ‘earth’ (other dialects *hagāzī*, *hāk*). Intervocally all dialects prefer to keep *-h-*, but e.g. *trahag* (frequently in Ra *traag*) with minimum hiatus. In pre-consonantal position after a short vowel it is dropped in all Ra, with compensatory lengthening: *ahwāl* ‘news’ > Ra *awāl*, *šahr* ‘town’ > Ra *šār*, *zahg* ‘child’ > Ra *zāg*. Finally *-h* is dropped in all Ra after a long vowel: *kō* ‘mountain’, *rā* ‘road’, otherwise it is kept in all dialects [kōh, rāh].

m [m] voiced bilabial nasal. In EHB *-m* > *-w-* in some words (chiefly *šwā* ‘you’, *hawān* ‘that’, *khawān* < NP *kamān*, other Balochi dialects *kamān* ‘bow’, *nēway* ‘half’, *nawāš* ‘prayer’, *nyāwāš* ‘middle’).

n [n] voiced post-dental nasal. It has three non-phonemic allophones: before retroflex *ʈ*, *ɖ* [ŋ]; before *č*, *ʝ* [ɲ]; before *k*, *g* [ŋ]. In all dialects except AfRa, Sa, La, vowel + *n* > *Ń*, except when a vowel follows, when *n* remains. In EHB the process has gone much further: *Ń* + *n* + consonant > *Ń* in all final positions. The tendency is also present in Ke, Co, and Sa, but to a much lesser extent. La appears to nasalise a final vowel + *n* and drop *-n*.

ɳ [ɲ] exists as a phoneme independent of *ɖ*, *ʈ* only in EHB, where it also can alternate with *r*. It does not occur initially.

[ŋ] is only an allophone of *n* before *k*, *g*—in all dialects. Note that *bāng* = /bāŋg/, *angūr* = /aŋgūr/, *tirōnk* = /tirōŋk/, etc.

l [l] voiced post-dental lateral, non-fricative. It tends to be medium clear.

r [r] voiced post-dental single flap; when doubled /rr/ it has three or more flaps (“trill”).

r [ɽ] voiced back-alveolar retroflex flap, does not occur initially. In some words in Ra it can be an allophone of *ɖ*: *gura* ‘then’, *abōr* ‘woe!’.

w [w] voiced labio-dental frictionless continuant, unrounded. In some Ra areas and in EHB it can be bilabial, almost [β]. But in EHB it is bilabial only occasionally post-vocally and before a consonant: /hiwda/ ‘17’ [hiβda]. It has been recorded as aspirated initially in the words *whāw* ‘sleep’, *whār* ‘food’, and *whāš* ‘glad’, though just as often these words lack any aspiration. As an epenthesis in Ra, see § 37.4.3.1.

y [j] palatal frictionless continuant. As an epenthesis in central dialects, see § 37.4.3.1. It is never aspirated initially in EHB.

37.5. Stress

Balochi has a fairly strong stress accent. A long vowel or diphthong attracts the stress, which is usually on the first of several long syllables. In words of several short syllables, stress is on the first syllable. Final *-ī* does not attract stress, except in the word *marōčī* [ma:ro:či:] ‘today’.

37.6. Geminates

All consonants can be geminated (lengthened). When not etymological or phonemic, purposive gemination of the final consonant of short-voweled monosyllables is a means of emphasis. In all such monosyllables the final consonant can be geminated, especially if the next word has an initial short vowel. Gemination is, however, also often phonemic, as in:

<i>čam</i> : <i>čamm</i>	‘trick’ : ‘eye’
<i>dik</i> : <i>dikk</i>	‘hillock’ : ‘advance’
<i>pur</i> : <i>purr</i>	‘cold ashes’ : ‘full’
<i>galag</i> : <i>gallag</i>	‘herd’ : ‘wheat’
<i>jal</i> : <i>jall</i>	‘fish net’ : ‘case, box’
<i>jan</i> : <i>jann</i>	‘wife’ : ‘wedding procession’
<i>jaṭ</i> : <i>jaṭṭ</i>	‘live ashes’ : ‘Jat tribesman’
<i>jat</i> : <i>jatt</i>	‘struck’ : ‘camel driver’
<i>gar</i> : <i>garr</i>	‘precipice’ : ‘mange’
<i>kad</i> : <i>kadd</i>	‘manure’ : ‘size’
<i>dur</i> : <i>durr</i>	‘far’ : ‘pearl’
<i>bir</i> : <i>birr</i>	‘lightning’ : ‘untamed’

A “balance law” dislikes a geminated consonant before a long vowel:

<i>garr</i> ‘roar’	1 sg. <i>garīn</i> , 3 sg. <i>garīt</i> ; but 3 pl. <i>garrant</i>
<i>girr</i> ‘pull along’	1 sg. <i>girīn</i> , 3 pl. <i>girrāt</i>

but:

<i>gir</i> ‘catch’	1 sg. <i>girīn</i> , 3 pl. <i>girāt</i>
<i>dir</i> ‘rip (intr.)’	1 sg. <i>dirīn</i> , 3 pl. <i>dirāt</i>

Ra *tunnag* ‘thirsty’, Co, La *tūnag* (*ū* by compensatory lengthening, *nn* > *n*).

Geminates are normally simplified in AfRa:

AfRa	Other dialects
<i>anūn</i>	<i>hannūn</i> ‘now’
<i>kaš-</i>	<i>kašš-</i> ‘pull’
<i>muč</i>	<i>mučč</i> ‘crowd’
<i>yak</i>	<i>yakk</i> ‘one’
<i>guṭ</i>	<i>guṭṭ</i> ‘throat’
<i>pisay</i>	South Ra <i>pissay</i> ‘father (g. sg.)’

37.7. Loanwords from Persian/Arabic

In native writing the original spelling of Arabic loanwords is usually kept, though some experiments have been made in phonemic spelling (principally by Gul Khan Nasir, 1914–1981, a leading poet). Pronunciation follows in the main Persian practice, except that

Ar *q* is [k]; but “elegant” [q].

Ar *θ*, *ḍ* are [s, z] also in EHB, but see § 37.4.4 under *θ*, *ḍ*.

‘*ayn* is lost in initial position; in word-medial position it becomes *h* (or where *h* is lost, a preceding vowel is compensatorily lengthened): *šāhir*, *šāir* < Ar *šāʿir* ‘poet’.

37.8. Metathesis

Sporadic, typically only in certain words; it is not as common as in Pashto:

<i>matbal</i>	‘meaning’, < <i>matlab</i>
<i>ḡagdal</i>	‘Jat language’, < <i>jadgal</i>
<i>-tk-</i>	p.p. (< * <i>kt</i> < * <i>xt</i>)
<i>rusxat</i>	‘leave’, < <i>ruxsat</i>
<i>bašk-</i>	EHB ‘give’, < <i>bakš-</i>
<i>trus</i>	‘fear’, < <i>turs</i>
<i>nēmṛōč</i>	all dialects (EHB <i>nēmṛōš</i>) ‘midday’; cf. also <i>nēmṛōč</i> all dialects, EHB <i>nēmṛōš</i>
<i>asr/l</i>	‘tear’ < <i>ar/ls</i> , all dialects
<i>dalwat</i>	‘wealth’, < <i>dawlat</i>
EHB <i>kulf</i>	‘key’, < <i>kuf</i>
EHB <i>kilba</i>	‘qibla’, < <i>kibla</i>

37.9. Lexicon

The Balochi vocabulary contains approximately

- 15% identified loanwords from IA
- 35% loanwords from Ar/NP
- 30% unknown: many are as yet unrecognized IA words
- 18% genuine Balochi (i.e. with a good Iranian etymology)
- 2% loanwords from Brahui (a few genuine loanwords from Brahui; otherwise words occurring in both languages without a convincing etymology)

Bibliography

General bibliographies of Balochi language studies are to be found in:

Compendium Linguarum Iranicarum s.v. “Balōčī.”

Encyclopaedia Iranica s.v. “Baluchistan. III. Language and Literature.”

*Encyclopaedia of Islam*² s.v. “Balūčistān. B. Language.”

See further:

Barker, M. A., and A. K. Mengal. 1969. *A Course in Baluchi*, 2 vols. Montreal: McGill University, Institute of Islamic Studies.

Elfenbein, Josef. 1960. “Balūčistān. B. Language.” *Encyclopaedia of Islam*, new ed., 1: 1006–1007.

———. 1966. *The Baluchi Language: A Dialectology with Texts*. London: Royal Asiatic Society (Monograph 27).

———. 1979. “Report on a Linguistic Mission to Helmand and Nīmrūz.” *Afghan Studies* 2: 39–44.

———. 1989a. “Balōčī.” In Schmitt 1989 § 4.1.2.6, pp. 350–62.

———. 1989b. “Baluchistan. III. Language and Literature.” *Encyclopaedia Iranica* 2: 633–44.

———. 1990. *An Anthology of Classical and Modern Balochi Literature*, 2 vols. Wiesbaden: Harrassowitz [with bibliography].

Jahani, C. 1989. *Standardization and Orthography in the Balochi Language*. *Acta Universitatis Upsaliensis* XX/X.

Schmitt, Rüdiger, ed. 1989. *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert.

Spooner, Brian J. 1989. “Baluchistan. I. Geography, History, Ethnography.” *Encyclopaedia Iranica* 2: 598–632.

Armenian Phonology

John A. C. Greppin
Cleveland State University

This chapter deals primarily with the historical phonology of Armenian, perhaps the most complex of the Indo-European languages in that respect; all transcriptions of Classical Armenian herein are transliterations of the script rather than phonemic interpretations.¹ There is, alas, very little on the synchronic phonology of the modern language, whether eastern or western dialect. In fact, there are only two significant rules for pronunciation of single words, and no rules for variation in sequences of words, other than such an obvious rule as the unlikelihood of having a voiced consonant adjacent to a voiceless consonant at word juncture or in compounds. There are discussions of phonology in the Armenian grammars by Meillet (1936) and Godel (1975), but they provide little of interest for synchronic phonology. The principal works in Armenian (Djahukian 1974, Arakelian, Khachaturian, and Eloyan 1979, Muradian 1982) which would touch upon this aspect deal almost entirely with phonetic rules. However, in the primer for Modern Eastern Armenian, by Fairbanks and Stevick (1958), we get some insights into the realization of that dialect, specifically the Yerevan dialect. Fairbanks and Stevick give the text in Armenian letters, and then give a realization of it in Latin letters that is not a mere transliteration, but a phonetic representation. It is not perfect; in places there are errors (the initial *e-* of the verb ‘to be’ is incorrectly shown with prepalatalization in the Yerevan dialect), and it is not systematic. The principal differences are as follows. There is a fairly regular devoicing after *ɾ*, *ɹ*, *l*, and *t*: *derjak* ‘tailor’ = [dɛɾɕak], *atjik* ‘young girl’ = [axɕik], but *lurj* ‘serious’ = [lurj]. Nasals will sometimes be involved in this:

1. AUTHOR’S NOTE: Abbreviations: Ag., Agulis; Arm., Armenian; CArm., Classical Armenian; CilArm, Cilician Armenian; Erz., Erzurum; Gk., Greek; Goth., Gothic; Grg., Georgian; Hitt., Hittite; IE, Indo-European; Kar., Karchevan; Lat., Latin; Megh., Meghri; MP, Middle Persian; OHG, Old High German; OIr., Old Irish; Per., Persian; Phl., Pahlavi; Skt., Sanskrit; Sog., Sogdian; Sv., Svedia; Syr., Syriac; Tb., Tiflis; Tig., Tgranakert; WArm., Western Armenian. Capital letters indicate cover symbols for phoneme classes: C = Consonant or *ts*, *dz*, *tʰs*; Č = č, j, ě; L = *l*, *t*; N = *m*, *n*; R = *ɾ*, *ɹ*; T = any dental. The chart of consonants was prepared by Peter T. Daniels.

xndir ‘question’ = [xəntir]. Contrarily, there is frequent voicing of a non-aspirated voiceless stop after an *n*: *kentronakan* ‘central’ = [kendronakan], but *naxəntrel* ‘repose’ = [naxəntrel]; and a somewhat random devoicing of final voiced stops: *mēj* ‘middle’ = [mēc̥], *hog* ‘concern’ = [hok], but *arag* ‘swift’ = [arag].

It is very difficult to say that modern Armenian lacks sentence sandhi, yet I have found only the tiniest amount (no more than I knew before I began this investigation). This will duly be reported in the appropriate place in this article. In the meantime, I urge phonologists to enjoy the rich and often unusual historical phonology of Armenian, as well as the variations found in the medieval and modern dialects.

The Armenian people are known historically from approximately 525 B.C., being mentioned in the Old Persian inscriptions of Darius. A further sparse mention of the Armenians is found in the *History* of Hecataeus of Miletus. The foremost Armenian historian of the Classical period, Movses Khorenatsi, provides a fabled early history for the preliterate Armenians which deals with the mythological age of heroes; he then manufactures Armenian contacts with the Babylonian and Assyrian people.

The Armenian alphabet dates from the earliest part of the fifth century, when the Armenians began to translate the Old and New Testaments. Hitherto the Armenians were dependent on both the Greek and Syriac churches for their liturgy. Yet the Armenians wanted episcopal independence, and this they could have only if the scripture and liturgy existed in their own language. Hence the Armenian alphabet and the beginning of Armenian literacy. This alphabet was devised under the leadership of Saint Mesrop Mashtots² and his efforts were recorded in a short but valuable biography, *Vark’ Mashtots’i* (Life of Mashtots), written by his disciple Koriwn. Shortly there was much literary activity, bringing the completion of the Bible, the writing of various histories, religious tracts and the translations of considerable patristic literature from Greek and Syriac. The Armenians thus greeted their new alphabet with a great burst of intellectual activity.

It seems clear that the language of the Armenian Bible is something of a koiné version, for there is reference to dialect diversity in the early authors. One fifth century author, Yeznik Koghbatsi, differentiates between his dialect and a “southern dialect”; other evidence for dialect diversity in the Clas-

2. It was largely necessary for the Armenians to have their own alphabet since existing alphabets did not approach the wide range of fricatives and affricates, voiced, voiceless and voiceless aspirate, that Armenian possessed.

sical period stems from the seventh and eighth century, especially from those Armenian grammarians who commented on the Armenian translation of *The Art of Grammar* by Dionysus Thrax. An anonymous commentary (Adontz 1915: 147.9) cites the dialect of Karch, and a later commentary by Stephen Siunetsi (Adontz 1915: 187.10) which cites additional dialects by name, among them his own, and the dialect of Artsakh (Nagorno Karabakh), a singular dialect now somewhat impenetrable to a standard Armenian speaker.

However, there is no basis for saying that these dialects were then considerably different from the norm, even the dialect of Artsakh. We may assume that all early Armenian speakers were capable of intercommunication with little difficulty. This state conforms with what we know about the ancient Greek dialects, for in spite of their recorded diversity, mutual communication was certainly not too difficult. Strong evidence for dialect diversity comes by the tenth century, by which time we have solid evidence for what we now call the Western Armenian dialects. Yet, we can agree with Josef Karst (1901) that there were well developed vernaculars from the fifth century.

There have been, over the years, three major systems of classification of the Armenian dialects. The first was espoused by Hrachia Adjarian, who saw three separate dialect groups, based on morphological principles. He noted (1909: 13–14) differences based on the formation the present tense.³ Somewhat later, Ararat Gharibian (1941) suggested a new arrangement, a fourfold grouping, based on phonological differences. Finally, Academician Gevork B. Djahukian formulated (1972) a multifeatured classification which maintains a twofold sectioning, Eastern and Western, but with considerable fine tuning: there are seven sub-groups for the Western dialects, and four for the Eastern.

In addition to the standard Classical Armenian, which remained the norm for six hundred years, there also flourished, during the sixth century, a dialect called Hellenized Armenian. This written dialect, wholly artificial, was originated by Armenian Hellenophiles who wished to translate Greek classical and patristic texts in a way that reflected Greek grammar. Greek words were borrowed when an easy Armenian equivalent was not available; even Greek prepositions were fair game. Armenian syntax was altered to conform to Greek syntax and we are left with writing that is often incomprehensible,

3. One group had its present tense end in *-um*, the second had its present tense begin with *gə*, and the third, surely a small group, based its present tense on the infinitive in *-el*.

or, at least, open to multiple interpretations. This artificial dialect never flourished and can be regarded as a hothouse literary aberration.

Classical Armenian is a language with an extraordinary abundance of loanword layers. The greatest seems to be the Iranian substratum which Émile Benveniste (1964) identified as Parthian. So large is this stratum that, before Hübschmann (1897) was able to show that Armenian was a separate IE language, Armenian was regarded as a renegade Iranian dialect. The Iranian stratum in Armenian has, in fact, more roots than the inherited IE stratum. For it Hübschmann cites 438 IE roots, but 686 Persian stems by loan. This stratum is evident in the earliest body of Armenian literature and more was added later, along with Arabic. Considerably earlier than this Iranian influence was a Hurro-Urartian invasion (Greppin 1992), and a Hittite influence, though the latter was more likely late Luwian.

There is also a body of Syriac and Greek vocabulary. Hübschmann notes 133 Syriac stems and 512 from Greek. To a very great extent, this vocabulary is derived from Christian religious texts that were translated into Armenian. From medieval times on there was continuing intrusion from Arabic and, later, from Turkish. After the twelfth century the Armenians slid into an ever increasing dark age, until by the eighteenth century, an imprecisely known but no doubt large portion of Armenian day-to-day vocabulary was foreign. But following a renaissance inspired by the Armenian Mekhitarist monastic order in the eighteenth century, this vocabulary was rooted out of standard speech, and today the spoken language, both Eastern and Western dialects, has reverted to a more pristine level. But during this period of strong foreign influence, new sounds became common and two letters, *ō* and *ḥ*, were added to the alphabet, about which more will be said shortly.

38.1. Vowels

Armenian has seven vowels: *u a ɛ ɪ ē ʁ ə* *i n o m u* and four diphthongs, *u j ay uɪ aw* now [ō] *ɛɪ ew* now [ev] *ɪ j oy*. All but shwa come from an IE source.⁴ All is orderly except for the reflex of **ow* and **ew* (> *oy*) as against **oy* and **ey* > *ē*. There is apparently a lack of symmetry here, for one would expect **ey* and **oy* to produce a diphthong analogous to the *oy* that comes **ew* and **ow*. It seems clear that Mesrop Mashtots, who devised the alphabet, regarded the reflex of the **-w* diphthongs in Armenian as monoph-

4. IE **a* and **ā* > Arm. *a*; **e* > *ɛ*; *ē* > *i*; **i*, **ī* > *ɪ*; **o* > *o* and *a*; *ō* > *u*; **u*, **ū* > *u*; **ay* > *ay*; **aw* > *aw*; **ey*, **oy* > *ē*, **ew*, **ow* > *oy*, **R* > *aR*, **N* > *an*.

thongs (*u*) while the reflex of the *-y diphthongs was apparently a monophthong (*ē*). There is scarce evidence that long *ē* is a vowel long in duration, and it is most puzzling why it is not written **ey* on analogy with Arm. *oy*.

There is another system in which Arm. *ē* and *oy* continue to be out of step. When an Armenian root with an *ē* or *oy* is then followed by another syllable, the original *ē* or *oy* will become an Arm. *i* or *u*. Examples are Arm. *gēt* ‘learned (person)’, *gitanem* ‘I know’ and Arm. *yoys* ‘hope’, *yusem* ‘I hope’. Correspondingly, original roots in *i* and *u* will be reduced to [ə].⁵ Examples are Arm. *lič* ‘lake’ and *bun* ‘nature’, which, with further suffixation, become *lčak* ‘pool’, *bnakan* ‘natural’. There is, of course, a vocalic gap between these two consonants enclosing the original short *i* or *u*, and it is pronounced as [ə]. Thus **CuC+*, **CiC+* > Arm. [CəC+]. In Middle Armenian, in some circumstances, this shwa is actually written in. This rule applies to Armenian words of any origin, IE or loan.

Godel (1975: 6), following Meillet (1936: 168), states that Armenian long *ē* and short *e* differ one from another not by length, but by timbre. Further, *ē* does not model itself after the other pure vowels (*a e i o u*).⁶ It seems that the distinction between *e* and *ē* vanished at least by the medieval period except in initial position, where Arm. *e* was pronounced [ye] and *ē* remained without initial homorganic glide. Now there is no Armenian dialect which differentiates, medially nor finally, between the two shapes of *e*, and it should be noted that loanwords from Greek with an η did not express that η in Armenian with a long Arm. *ē*, but rather η and ε were united and usually transliterated with a simple Arm. *e*.

The lack of an early Arm. **ō* in apposition to Arm. *ē* further adds to the problem of lack of symmetry since the transliteration of Greek ω in loanwords does not parallel *ē*. Rather, a Gk. ω is transliterated as Arm. *ov*. Here note Gk. ὠκεανός = Arm. *ovkeanos*, Gk. ὄσαννά = Arm. *ovsannay*, and the Armenian proper name *Movsēs* = Μωσῆς.

In spite of the fact that there is no difference in timbre or length between Arm. short *o* and *ō*, they do differ in initial position where, paralleling Arm. *ē* [e] and *e* [ye], short *o* is pronounced with an initial glide [vo]; long *ō* is simply [o].

5. There are regular exceptions: monosyllables ending in *-i* and *-u* will keep the vowel when a suffix is added: Arm. *ji* ‘horse’, *jiawor* ‘rider’; Arm. *bu* ‘owl’, *buēč* ‘a horned owl’.

6. Arm. *a e o* are called “stable” vowels, while *i* and *u* are “unstable,” passing to [ə] when a new syllable is added to the root within which they are.

The question of the earliest pronunciation of Arm. initial *o-* is perplexing. Though there is ambiguous evidence, it seems more likely that Arm. *o* had a homorganic glide (= *vo*) from the earliest literate times.⁷ This cannot be tested with IE words since every IE **wV-* came into Armenian as *gV-* (Arm. *gini* ‘wine’). Secondly, there are no examples of Syriac loans with an initial Armenian *v-*. There are, however, two Armenian words that are spelled initially with the letters *v* and *o*, which would seem to reveal a pleonasm. One is *vohmak* ‘a herd’, known from the fifth century in Chrysostom, but it has an alternate spelling of *yohmak* as well as *vomak*, and it would seem also that this word has suffered some contamination. To this may be added Gk. βόρεας ‘north wind’, brought into Armenian with the spelling *vorē* ‘id.’. However, there are also alternate spellings *vōrēas*, *vawīēk*. Further, this word first appears in the Armenian translation of Aristotle, and probably was a example of phonemic hypercorrection, a product of the sixth century Hellenizing school.

The same exercise can be repeated for Arm. *e*, which was pronounced as [ye] when initial. When a simple [e], without *y*-glide, was needed initially in loanwords, the letter *ē* was employed.

There are peculiarities involving the vowels *u* and *aw*. The former is written as the digraph *ow*, and the latter, a digraph, which was soon replaced by long *ō*. One suspects that the digraphic spelling of Arm. *u* = *ow* is an orthographic calque on Gk. *ov*. Certainly Arm. *ow* is a monophthong now, and no transcription text shows *ow* as a diphthong. Similarly, historical Arm. *aw* is now always written as a monophthong *ō*.

More complex is the pronunciation of Arm. *oy* and *ay*. Except when word final, they are always the equivalent of a vowel followed by a *y*-glide. When final, *-ay* and *-oy* are pronounced as *-a* and *-o*. There is a good explanation for this phenomenon, because words otherwise ending in a vowel will develop a *y* extension when another morpheme, bound or free, is added. Thus, such an ancient loanword as Gk. ἀπόρρυμα ‘a type of Greek measure’ was transliterated in Armenian as *aporiumay*,⁸ presumably with a silent final *-y*, but in oblique cases the *y* will sound: instrumental *aporiumayaw*; similarly, such a modern word as ‘California’, to which the locative in *-um* is added, will become *K'aliforniayum* ‘in California’. The same rules apply to a word

7. This glide came to occur only initially, for loanwords into Armenian with a medial *-vo-* always appear with the Armenian letters *-v* and *o-*: e.g., Gk. ὁμοσιτής (with Asiatic Greek pronunciation) = Arm. *avoristēs* ‘companion’.

8. This transcription of Gk. /u/ by Arm. /iu/ would support the view that aspiration of initial Gk. /u/ results from an earlier PGk. ***u* > **yu* > *û*; thus **upo* > **yupo* > *ûπό*.

ending in *-oy*. It thus appears that the digraphs *-ay* and *-oy* are orthographically correct; there is simply a secondary rule that says that when word final, *-y* is not pronounced.

38.2. Consonants

See Table 38-1.

Table 38-1. Classical Armenian Consonants

	Labial	Dental	Dental Affricate	Palatal	Velar	Laryngeal
Plosives						
Voiceless	p պ	t տ	c ժ	č ճ	k կ	
Aspirated	p' փ	t' թ	c' ց	č' շ	k' ք	
Voiced	b բ	d դ	j ձ	ǰ ղ	g գ	
Fricatives						
Voiceless	f ֆ	s ս		š շ	x ք	h հ
Voiced	v վ	z զ		ž ժ	ḡ ց	
Nasals	m մ	n ն			ŋ	
Laterals		l լ			ɭ լ	
Resonants		r ը		ɾ ռ		
Glides	w լ			y յ		

38.2.1. The Indo-European stops

Armenian consonantism differs considerably from those of the other IE languages. The language maintains a series of three similar phonemes: voiced, voiceless, and voiceless aspirated, and it has *satəm* characteristics as well. Thus IE **p*, **t*, **k*^w, **k*^ʷ > Arm. *h* or zero (Arm. *het* 'footprint, trace', and Arm. *otn* 'foot', Lat. *ped-* 'foot'), *t*^ʰ, *k*^ʰ, *s*; IE **b*, **d*, **g*^w, **g*^ʷ > Arm. *p*, *t*, *g*, *c* (= [ts]); IE **bh*, **dh*, **g*^w*h*, **g*^ʷ*h* > Arm. *b*, *d*, *g*, *j* (= [dz]). There are also two affricate series, dental and palatal, which are derived from clusters: Arm. *c*^ʰ (= [ts^ʰ]), *č*^ʰ; *c* (= [ts]), *č*; and *j* (= [dz]), *ǰ*.⁹ In addition there are two sibilants, *s* and *š*, two laterals, *l* and *ɭ*, and two resonants, *r* and *ɾ*. These varied series

9. These develop from PIE in the same way that the Germanic consonants developed from PIE, in the cycle *Media* > *Tenuēs* > *Aspirata* > *Media*.

could imply that the Armenians had been long settled in Western Asia adjacent to the languages of Daghestan and began, through proximity, to conform to the consonantal richness we might assume for those languages in the first millennium and which is now plenteously abundant.

38.2.2. *Affricates*

The palatal affricates *č*, *č'*, and *ǰ* are known in Armenian words derived from the large, unknown, pre-IE substratum, and in words of IE origin, from IE **Ty*-¹⁰ and K**y*-.¹¹ But there are problems, for there exists some evidence that **Ty*, in addition to producing a palatal affricate *Č*, might also produces the dental affricate *C* (= [TS]). I, however, feel that this proposed sound shift of IE **Ty*- to Arm. *C* is unlikely. The series is elliptical—there are no examples of **ty*- passing to Arm. *c* and the single existing etymology for **dhy* > Arm. *j* is semantically weak, and of the half a dozen possible etymologies for the passage of **dy* to Arm. *c*, all are, in some small way, flawed.

38.2.3. *Sibilants*

The reflexes of IE **s* also present a problem. It is clear that, before a vowel, the results are both Arm. *h* and zero: Arm. *hin* 'old', Lat. *sen*; Arm. *at*, Lat. *sal* 'salt'. By and large the rules states that IE **s*- gives Arm. *h*- before IE **e* or **i*, and zero elsewhere, yet there are aberrations.

The origin of *š*, from an IE base, is obscure, and there are precious few examples which may be summoned. A rather straightforward example exists in Arm. *ēš* 'donkey', Gk. ἵππος, Skt. *āśva*-, IE **ekwo*-. And clearly, Arm. *šun* 'dog' is related to Lat. *canis*, Skt. *śvā*, IE **kwōn*-. Yet there also exists Arm. *skund* 'pup', Goth. *hunds*, which is also reducible to IE **kwōn*-*t*-. This seems to leave us the apparent evidence that IE **kw* produces, twofoldly from the same root, Arm. *šu* and *sk*. Further, the *š* of Arm. *šun* and *ēš* cannot be shown to be a result of the following *u*, for we note Arm. *surb* 'holy', Skt. *śubhrā*-, IE *kubhra*-.

There have been varied efforts to show that IE **gh* and **dh* can yield Arm. *z*. Arm. *bazum* 'much' is often cited as an example, noting Skt. *bahú* 'id.'. Yet, acknowledging Hitt. *panku*, the IE form must be **bhng̃hu*- which would yield Arm. *banzum*. Efforts from IE *dh* are similarly fruitless. Arm. *z* cannot be

10. Examples are Arm. *mēj* 'middle', IE **medh*-y- (Lat. *medius*); Arm. *oročem* 'chew, gnaw', IE **rod*-y- (Lat. *rodo* 'gnaw'); Arm. *koč'em* 'call', IE **g^hot*-y- (Goth. *qīpan* 'say').

11. Examples are Arm. *jil* 'nerve, sinew', Lith. *gýsla*. IE **g^hhisl*-; Arm. *goč'em* 'cry out', Skt. *vakti*, IE **wok*-y-. I have no valid example for Arm. **č*. Note also that *n*-y and *r*-y produce *nj* and *rj*: Arm. *anutj* 'dream', Gk. ὄνειρος 'id.'; Arm. *sterj* 'barren', Lat. *sterilis* 'id.'.

derived from any IE source, but comes by loan. Note Arm. *zūr* ‘useless, vain’, Per. *zur* ‘false, lie’; Arm. *zep’iwr*¹² ‘Westwind’, Gk. ζέφυρος; Arm. *zopay* ‘hyssop’, Syr. *zōpā*.

Arm. *ž* does not result from any IE cluster, but is known from loanwords: Arm. *žam* ‘time’, Per. *zamān*, *žamān*.

38.2.4. Laterals

We know from our earliest loanwords that Arm. *l* would replace Persian and Syriac¹³ while Gk. *λ* was always replaced by Arm. barred *l*,¹⁴ presumably a dark *l*. Later this barred *l* developed a phonetic resolution approaching a voiced velar fricative, *g*, and became the equivalent of Persian *q*; hence Arm. *Łarabat* ‘Qarabaq’ (= Karabakh).

IE **l* by and large produces Arm. *l*, but when final after a consonant (-C/) it becomes *l̥* (*astl̥* ‘star’); when preconsonantal in medial position (*otb* ‘lament’); and when intervocalic (*kat̥in* ‘accord’). When sonant, we get Arm. *aL*: Arm. *gatt* ‘secretly’, Lith. *-vilti* ‘deceive’.

38.2.5. The Indo-European glides

Initial **y-* is a problem. It has been argued to pass to Arm. *l-* (*luc*) ‘yoke’; or *j* (*janam* ‘try’, Gk. ζῆλος ‘zeal’). There is also evidence for initial **y-* > Arm. zero: Skt. *yas* ‘who’, Arm. *or* ‘who [interrogative]’, Arm. *ur* ‘where’, *ok* ‘someone’, all from IE **yo-*, the relative pronoun; and note also Arm. *nēr* ‘sister-in-law’, Gk. εινώτερες, ‘id.’, IE **yenH-ter* > Arm. **hineyr̥* > **hnēr* > Arm. *nēr*. To support further the movement of **y* to Arm. *l*, we can also note Arm. *leard* ‘liver’, Skt. *yákṛt*. Some laryngeal solutions were tried. IE **Hy* was to produce Arm. *j̥*, unaffected **y* would continue as zero. Thus the parallels between Arm. *janam* ‘try’, Gk. ζῆλος ‘zeal, eager rivalry’, Skt. *yátar* ‘avenger’, IE **HyeA-*; also Arm. *jov* ‘sprout’, Gk. ζειαί ‘spelt’, Skt. *yávasa-* ‘grass’, IE **Hyow-*. And though the parallelism between Gk. *ζ-*, Arm. *j̥-*, and IE **Hy-* is interesting, and consistent with some views on laryngealism, the evidence is sparse.

IE **w* regularly passes to Arm. *g* before a vowel: Arm. *gitem* ‘know’, Lat. *video* ‘see’; Arm. *aganim* ‘spend the night’, Skt. *vásati* ‘live, dwell’, Gk. ἄ(ϝ)εσα (νύκτα) ‘spent (the night)’. Preconsonantal **w* remains: Arm. *awjin* ‘necklace’, Gk. ἀγχήω ‘neck’.

12. Arm. *p*’ is the consistent transliteration of Gk. *φ*.

13. Note: Syr. *ēlūl*, Arm. *ilul* ‘Hebrew month name’, Per. *lakht* ‘club, mace’, Arm. *laxt* ‘id.’.

14. Gk. λεπτόν, Arm. *tepton* ‘a small coin’.

There is a very unusual shift in which IE **dw-* becomes *erk-*. The most commonly cited example is IE **dwō* ‘two’, Arm. *erku*. Other examples include Arm. *erknč’im* ‘to fear’, Gk. *δραείδω* ‘id.’, Arm. *erkar* ‘long’, Gk. *δηρός* < **δᾱῶ-ρος* ‘id.’. The sequence probably is **dw-* > **dg-* > **tg-* > **tk-* > **θk-* > **-rk-* (*r* cannot occur initially) > *erk-*.

38.2.6. *Resonants*

There are two resonants, Arm. *r* and *i*. In loanwords, Gk. *ρ* is reflected as both Arm. plain *r* and dark *i*, a trilled and palatalized resonant. Hence Arm. *rabbi*, Gk. *ράββι*; Arm. *kiwī*, Gk. *κύριος* ‘lord’; Arm. *ropē*, Gk. *ρόπη* ‘moment’; Arm. *lambiurint’os*, Gk. *λαβύρινθος* ‘labyrinth’. Hübschmann (1897: 328) attributes the difference to the time when the words were received, the palatalized [r] being found in the earliest texts translated from Greek while plain [r] is known from later texts. There are a few instances of initial Arm. *hī*: Arm. *hīog* ‘pay’, Gk. *ρόγα* ‘stipend’, but such instances might be hypercorrections. Note also *hīrom* ‘Rome’. When sonant, IE **r* is reflected as Arm. *ar*: Arm. *arbi* ‘drank’, Lat. *sorbeo*. Syriac /r/ appears as Armenian *i* when initial (Arm. *iabunay*, Syr. *rahbūnā* ‘earnest money’) and flat *r* elsewhere (Arm. *karkureay*, Syr. *qarqūrā* ‘boat’). The reflexes of Persian /r/ in Armenian are similar. When initial it is realized as Arm. *i* (Arm. *iṛah* ‘road’, Per. *rāh*), elsewhere it is *r* (Arm. *hṇar* ‘means, way’, Per. *hunar* ‘craft, skill’). IE **r* comes through both as *r* and *i*, the latter in contact with *s*: note Arm. *or* ‘anus’, OHG *ars*.

38.2.7. *Indo-European nasals*

IE **m* and **n* are transferred without damage to Armenian except when vocalic. When sonant they are reflected as *aN*. Note Arm. *k’šan* ‘twenty’, Lat. *viginti*; and the Arm. negative particle *an-*; for **ṁ*, see Arm. *amain* ‘summer’, OHG *sumar*, IE **sṁar*.

38.2.8. *Velar fricative*

This common consonant, Arm. *x*, can appear in Armenian words which correspond to words in Hittite. Thus Arm. *xotxotem* ‘slaughter’, Hitt. *hul-huliya-* ‘fight, struggle’; Arm. *xat* ‘song, game’, Hitt. *halliyari-* ‘musician’. If these terms are related, then they must be so related by loan, for it seems fairly clear that these particular terms shared by Hittite and Armenian are not of IE origin. An Arm. *x* can also be derived by loan from Persian and

Syriac. Note Arm. *xarak* ‘rock, stone’, Per. *xārā* ‘id.’; Arm. *xarak* ‘sword’, Syr. *ḫarbā*.¹⁵

38.2.9. Aspirated consonants

Armenian aspiration, of IE origin, regularly appears after *t*, *k* when they reflect IE **t* and **k*.¹⁶ Arm. *h*- also appears in many words of IE origin which appear with an *h*- in Hittite. Thus Arm. *haw* ‘grandfather’, Hitt. *huhhas* ‘old man’; Arm. *hatanem*, Hitt. *hattai*- ‘pierce, stab’ are correspondences of which there are many, and which would seem to logically reflect a pre-vocalic laryngeal in IE. This view is still contested.

As to the laryngeal being a source of aspiration in Arm. *p’*, *t’*, *k’*, it seems that there are very few solid etymologies representing core vocabulary that would support such a view. The derivation of Arm. *t’* from IE **tH*, as in Arm. *ort* ‘calf’, Skt. *pr̥thuka*- from IE **pr̥tH*- is unnecessary since IE **t* would regularly produce that without a supplementary **H*. Note Arm. *t’anjr* ‘thick’, Lith. *tánkus* ‘id.’. On the matter of laryngeal aspiration in Armenian, see Greppin 1984.

38.2.10. Miscellaneous consonants

Two phonemes remain, Arm. *ֆ* *f* and *օ* *ō*. The former letter was put into the alphabet by the tenth century to represent Persian and Arabic words with *f*.¹⁷ And Arm. *ō* was used to replace a foreign initial *o*, which had hitherto been replaced by Arm. *ո* *o*,¹⁸ but this came to be pronounced [vo]: hence the need for an initial [o-], and thus Arm. *օ* *ō* (so written). Others argue that the original diphthong /aw/ became a monophthong by the tenth century, and merged with the recently borrowed sound /ō/.

38.3. Middle Armenian dialectology

38.3.1. Cilician Armenian

Written material in this dialect covers the period approximately from the tenth through the thirteenth centuries and is well described by Josef Karst

15. Gk. *χ* produces Arm. *k’*: Arm. *k’aradr* ‘plover’, Gk. *χαράδριός* ‘id.’.

16. The reflex of **p*, Arm. *h* or zero, is described above.

17. Hitherto a loaned *f* had been replaced by an Arm. *p’*; as also Gk. *χ* and *θ* were replaced by Arm. *k’* and *t’*. Here note Arm. *t’ermon* ‘caudle’, Gk. *θέρμων* ‘warm’; Arm. *p’alkon* ‘falcon’, Gk. *φάλκων*. For Arm. **k’*; see reference to Arm. *k’aradr* above.

18. In the earliest period, long foreign *ō* was replaced by Arm. *ով* *ov*: note *ovkianos* ‘ocean’, Arm. *Movsēs* ‘Moses’.

(1901). It maintains features that are parallel to the principal changes of the Western Armenian dialects, but also includes the voicing of velar fricatives: Arm. *sxal* 'wrong' > CilArm. *stal* (*t* = [g]). In addition, Karst describes certain combinatory changes. CArm. *-nr* > *-ndr*: CArm. *canr* 'heavy' > CilArm. *candr*; CArm. *mr* > CilArm. *mbr*: CArm. *armaw* 'date (fruit)' > *amraw* > CilArm. *ambraw*. There is also some metathesis with resonants. CArm. *datark* 'empty' > CilArm. *tardag*. Similar metathesis is noted in CArm. *sxtor* 'garlic', which is realized as CilArm. *xstor*. Resonants may also drop before a stop: CArm. *surb* 'holy', CilArm. *sup*. Medial syncope is also common: CArm. *t'agavor* 'king', CilArm. *t'akvor*.

38.4. Modern Armenian dialectology

38.4.1. Consonants

Earlier I mentioned the various systems for classifying the Armenian dialects, the most accurate system being the two-part division into Eastern and Western dialects according to phonological principles. This will be the pattern that will be assumed for this discussion of the dialects.¹⁹

The major phonetic difference between Eastern and Western dialects is a reversing of the consonantism. The Eastern dialect is more conservative, and maintains the Classical Armenian tripartite differentiation T T^h (written T^{*}) D. Thus, the reference pattern for the Eastern dialects is [p'–p–b] [c'–c–j] [t'–t–d] [č'–č–j] [k'–k–g]. A most clear example is found in the dialect of Meghri; there we have CArm. *dadar* 'pause' > Megh. *dedar*, CArm. *tapak* 'frying pan' > Megh. *tapak*, CArm. *č'ap'el* 'to measure' > Megh. *č'ep'il*. But all the eastern dialects are not so faithful to this symmetry. The dialect of Tiflis has a more diverse development of the voiced and voiceless non-aspirated stops. For although D usually remains as D, it can also become T^{*} medially.²⁰ And Van, which has the morphology of the Western dialects, develops its consonantism in a unique way, and is closer to the Eastern dialects in that respect. It produces from CArm. D T T^{*} a twofold T T^{*}.²¹

The canonical form for Western dialect development is also tripartite, with CArm. T D T^{*} becoming WArm. D T T^{*}, a pattern noted clearly in the

19. For easy access to a survey of the major dialects, see Greppin and Khachaturian (1986), where one can find further bibliographical information.

20. CArm. *bari* 'good', > Tb. *bari* but CArm. *xabel* 'to cheat' > Tb. *xap'il*.

21. CArm. *aygi* 'garden, vineyard', > Van *ikyî*; CArm. *kap* 'knot' > Van *kap*; CArm. *k'ani* 'how' > Van *k'yäni* (*ä* = [æ]).

dialect of Svedia.²² However, the Western pattern is largely bipartite²³ with the voiceless aspirate T̥ and voiced D merging as T̥ and Classical Armenian T becoming Western D. Thus Classical Armenian [k'–k–g] becomes Western dialect [k'–g–k']. Accordingly, Classical Armenian appears as follows in the (Western) dialect of Tigranakert. CArm. *getin* 'earth' > Tig. *k'edin*, CArm. *kēs* 'half' > Tig. *ges*, CArm. *k'atak'* 'city' > Tig. *k'ātāk'*. The other stops: labial, palatal, dental, etc., all behave accordingly.

But there are numerous variations. The (Eastern) dialect of Agulis reflects an original voiced stop D as both Ag. D and T̥. And though a voiced stop commonly remains voiced (CArm. *jor* 'valley' > Ag. *jiwr*), as expected, it will become a voiceless aspirate after a nasal or resonant (CArm. *erg* 'song' > Ag. *yerk'*), a continuing acknowledgment of the role of nasals and mutes on a following stop in the Armenian dialects.

There is also some evidence for secondary palatalization. In Karchevan we note CArm. *ink'n* 'himself' > Kar. *ink'i*, CArm. *girk* 'book' > Kar. *g'ürg'*; CArm. *keank'* 'life' > Meghri *g'ānk'*, CArm. *bambak* 'cotton' > Shamakhi *pāmpäg'*, CArm. *gomēš* 'buffalo' > Shamakhi *g'ümüş*. These are all scattered instances and there seems to be no observable pattern.

In the dialect of Erzurum (Karin) we note initial pharyngealization. Thus CArm. *uriš* 'other' > Erz. *huriš*, and CArm. *akn* > Erz. *hag*, but with such other features as palatalization or prothetic aspiration, there is no discernible pattern.

In some instances it would appear that there is a movement of *h* > *f* or *v*. Thus we have CArm. *hor* 'son-in-law' > Lori *fwor*, CArm. *hot* 'odor' > Karchevan *væt*, Goris *vot*, Haĭin *foed*. This shift, however, more likely came about from original *ho-* becoming deaspirated, followed by a prothetic homorganic glide *v* which later became devoiced (thus *ho-* > *o-* > *vo-* > *fō*). Other curious shifts include aspiration of voiced consonants (CArm. *botk* 'horseradish' > Haĭin *b'uetg*; CArm. *borot* 'wasp' > Mush *b'orwod*). There is also a continuing problem of unwarranted initial aspiration. This began with early, preliterate (note CArm. *Hoktemberi* 'October') and might occur in words of IE origin, though the mechanisms of initial aspiration of words inherited from IE are still much debated; there is later secondary aspiration

22. CArm. *pag* 'kiss' > Sv. *bək*, CArm. *p'or* 'belly', Sv. *p'är*.

23. This pattern is not wholly supported by the studies made of the Western Armenian dialect speakers who came for refuge to Soviet Armenia following the massacres in 1915–1922. In many instances we note that CArm. D passing to both D̥ and T̥ and sometimes T. Now, however, it is very unusual to find a Western dialect speaker, in diaspora, who keeps a three-fold division of stops.

as well (CArm. *aganim* ‘spend the night’ > Mush *hak’nil*, CArm. *alewor* ‘gray-haired’ > Mush *halvor*). Additionally, we might note a regular passage of *h* > *x* in the dialects of New Julfa and Van (CArm. *hayr* ‘father’, New Julfa *xer*; CArm. *mah* ‘death’ > Van *max*. This shift, regular in the dialects of New Julfa and Van, occurs spasmodically elsewhere; note CArm. *pah* ‘guard’ > Hajin *box*, CArm. *hayr* ‘father’ > Khoy *xer*).

38.4.2. Vowels

The vowels of the Armenian dialects are as varied as the constellations of the heavens. The original short vowel system of *a e i o u* can become almost anything. In Van dialect, for instance, the phoneme /a/ can become *a ä i ye o eo u iw e ə* and zero. There is also what seems to be the initial phases of a vowel harmony system developing in certain Eastern dialects (note Greppin 1985). It is clear, though, that the process is not yet fully developed, for there are frequent gaps in the system. This curious system surely results from the proximity of Azeri speakers. However, other such systems exist that are not influenced by Azeri, and here one might note a system of fronting in the dialect of Akn, a Western dialect, where the vowels *o* and *u* become *ö* and *ü* after dental stops (here see also Vaux 1993). This system is similar to that described in Adjarian’s Law (here see Muradian 1986, where /a/ > /ä/ after *b d g j ĵ* in certain central dialects).

38.4.3. Modern phonology

Allophonic variations in the standard eastern literary dialect are surprisingly few.

The most common rule shows devoicing and aspiration of a stop after *l*, *t*, *m*, *n*, and *r*; *i*. Thus Arm. *andam* ‘limb’ (< Phl. *andām* ‘id.’) is realized as /ant’am/; *surj* ‘round’ is /šurč/. This shift is similar to an ancient phenomenon where we can see spelling variations in final NC sequences from the fifth century; note Arm. *amb/p* ‘cloud’ (< IE **mbhos*), more commonly spelled *amp* (Arm. *amb* is expected according to the normal IE development laws). A similar devoicing also appears, with no deducible pattern, in medial positions. Thus Arm. *t’agawor* ‘king’ is now heard, in the eastern dialects, as [t’akavor].

The passage of CArm. initial *y-* to /h/ is universal in the modern dialects and known in the medieval Cilician dialect. Thus we have CArm. *yet* ‘after’, CilArm. *het* (Karst 1901: 34–35). Modern eastern orthography has been changed to accommodate this phenomenon, and an original prevocalic *y-* is now always realized as *h-* in the Soviet orthography (which is still in use).

The most complex phonological system deals with shwa insertion. Thus initial *sC* becomes /ə*sC*/: Arm. *sk'anč'eli* 'wonderful' is rendered as /ə*sk'anč'eli*/. This phenomenon is known on an earlier level, recognized in Armenian adjustments to Middle Persian vocabulary. Note Arm. *išxan* 'prince', which presumably came from a Parthian equivalent of OIr. **xšāna-*, Sog. *ḡš'wn*, and in spite of MP *šāh*; a parallel metathesis exists in Arm. *bdeasx* 'proconsul, mayor', Phl. *pātaxšāh* (note unmetathesized Grg. *pitiaxši*) 'id.'.

Similar is the morphophoneme {*z-*} which in Classical and Middle Armenian marks a definite direct object: thus *tesanem zmart* 'I see the man', as opposed to *tesanem mard* 'I see a man'). When the definite direct object marker was attached to a consonant-initial direct object, the *z-* took prothesis [ə*z-*]: [tesanem ə*z*mart].

As noted earlier, a vowel /*u*/ and /*i*/ in a sequence /*CiC*/ or /*CuC*/ will pass to zero when a suffix is added: thus *surč* 'coffee' but *srč'aran* 'coffeehouse'. There is then, necessarily, shwa insertion: /*sərč'aran*/. This same phenomenon appears in circumstances where the vocalized consonant is not a liquid or a nasal. Arm. *k'shan* /*k'əšan*/ 'twenty' (= Lat. *viginti*) < *gišan*.²⁴ A similar phenomenon occurs in **CR* sequences. Note *grem* 'I write' is pronounced as /*gərem*/; Arm. *blur* 'hill' > /*bəlur*/.

This rule is just about ubiquitous, functioning wherever there is a *CC* cluster, producing from it either /ə*CC*/ or /*CəC*/. Note the realization of the following proper name: *Mkrtč'yan* = [məkərč'yan < **mi/u-ki/urt-ič'-yan*].

I would close with a word on the phenomenon of ejectives in Modern Armenian. Comrie (1981: 180) stated that the Armenian voiceless unaspirated consonants (*p*, *t*, *k*, *c*, *č*) may be accompanied by a glottal catch. This point has further been noted in Armenian publications as well. Yet the distribution and frequency of Armenian glottalization remains unclear. It is noted, though, only in the speech of those from the Republic of Armenia, not elsewhere in the Eastern dialect (e.g., in Iran) and not in any of the Western dialects. Its presence in Eastern Armenian, a new phenomenon, surely shows the force of areal linguistics as well as the effect of time and proximity as a leveler of linguistic diversity. I think it is also safe to say that the once significant Armenian dialect of Tiflis, largely extinct since the Second World War, was never noted as a carrier of this phenomenon.

24. We note here the secondary rule that produces from a borrowed *ks* the Armenian realization with /*k'*/. Note also Arm. *k'sip'iē* 'swordfish', Gk. ξιφίας 'id.'; also MP *dušwarr* 'unfortunate', Arm. *t'suair* (which has an alternate spelling *č'uaiṛ*).

Bibliography

- Adjarian, Hratchia. 1909. *Classification des dialects arméniens*. Paris: Champion.
- Adontz, Nicolas. 1915. *Дионисий Фракийский и Армянские тоякобатели*. St. Petersburg: Press of the Imperial Academy of Sciences. French translation *Denys de thrace et les commentateurs arméniens* by . Louvain: Imprimerie orientale, 1970.
- Arakelian, Varag D., Amalya H. Khachaturian, and Seda A. Eloyan. 1979. *Žamanakakic' Hayoc' lezu*, hator 1: *hnč'yunabanut'yun ev baragitut'yun* (The modern Armenian language, vol. 1: Phonetics and lexicology). Yerevan: Press of the Armenian Academy of Sciences.
- Benveniste, Émile. 1964. "Éléments parthes en arménien." *Revue des études arméniennes* 1: 1–39.
- Comrie, Bernard. 1981. *Languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Djahukian, Gevork B. 1972. *Hay barbařagitut'yan neracut'yun* (Introduction to Armenian dialectology). Yerevan: Press of the Armenian Academy of Sciences
- . 1974. *Žamanakakic' hayereni tesut'yan himunk'nerə* (General theory of the Armenian language). Yerevan: Press of the Armenian Academy of Sciences.
- Fairbanks, Gordon H., and Earl W. Stevick. 1958 *Spoken East Armenian*. New York: American Council of Learned Societies.
- Gharibian, Ararat. 1941. *Hay barbařagitut'un* (Armenian dialectology). Yerevan: Press of the Armenian Academy of Sciences.
- Godel, Robert. 1975. *An Introduction to the Study of Classical Armenian*. Wiesbaden: Reichert.
- Greppin, John A. C. 1984. "The Reflexes of the IE 'Voiceless Aspirates' in Armenian." In *International Symposium on Armenian Linguistics, Yerevan, September 21–25, 1982: Reports*, 34–48. Yerevan: Press of the Armenian Academy of Sciences.
- . 1985. "A Note on Some Armenian Dialects Showing Umlaut." *Zeitschrift für vergleichende Sprachwissenschaft* 98: 245–49.
- . 1992. "Some Effects of the Hurro-Urartian People and Their Languages upon the Earliest Armenians." *Journal of the American Oriental Society* 111: 720–30.
- Greppin, John A.C., and Amalya H. Khachaturian. 1986. *A Handbook of Armenian Dialectology*. Delmar, N.Y.: Caravan Books.

- Hübschmann, Heinrich. 1897. *Armenische Grammatik*. Leipzig: Breitkopf und Härtel. Repr. Hildesheim: Olms, 1962.
- Karst, Josef. 1915. *Historische Grammatik des Kilikisch-Armenischen*. Strasbourg: Trübner. Repr. Berlin: de Gruyter, 1970.
- Meillet, Antoine. 1936. *Esquisse d'une grammaire comparée de l'arménien classique*. Vienna: Mekhitharistes.
- Muradian, Hovhannes D. 1982. *Hayoc' lezvi patmakan k'erakanut'yun*, hator 1: *Hnč'unabanut'yun* (A grammatical history of the Armenian language, vol. 1: Phonetica). Yerevan: Press of the Armenian Academy of Sciences.
- . 1986. "Excursus." In Greppin and Khachaturian 1986: 27–33.
- Vaux, Bert. 1993. "Coronal Fronting in the Akn Dialect of Armenian." *Annual of Armenian Linguistics* 14: 15–30.

This page intentionally left blank.

3 Dravidian Languages

This page intentionally left blank.

Brahui Phonology

Josef Elfenbein

Johannes Gutenberg–Universität Mainz

39.1. Introduction

Brāhūī (properly Brāhōī, conventionally Brahui) is the principal language of most Brahui tribesmen, who live mainly in a continuous territory in Pakistan, which extends from just north of the Arabian seacoast at approximately Las Bela northwards nearly to Quetta, and thence northwestwards to Nushki and across into Afghanistan along the Hilmand River as far as Sīstān; in Afghanistan they are not distinguished ethnically from the Baloch, with whom they share a habitat. Many Brahuīs also live in Karachi and in other large cities in Sind, as well as in Quetta.

It is to be noted that the terms “Brahui tribesman” and “Brahui speaker” are by no means coterminous: for many centuries Brahuīs have described themselves, especially to outsiders, as Baloch, particularly since the establishment of the Brahui Confederacy (“Khanate of Kalat”) in the 16th century (cf. Elfenbein 1987). At least two-thirds of Brahuīs are “bilaterally bilingual” (using Emeneau’s terminology) in Balochi and Brahui, and many Brahuīs do not speak Brahui at all (on these matters see Elfenbein 1990, especially pp. 434, 437–38). The number of speakers is probably of the order of 400,000 (350,000 in Pakistani Baluchistan or Sind, and 50,000 in Afghanistan). About two thirds, or 275,000, speak both Balochi and Brahui as mother tongues, and perhaps at most 50,000 speak only Brahui.

Depending on their habitats, migration routes, and social contacts, the second or third language spoken by Brahuīs can be Balochi, Urdu, Persian, Sindhi-Siraiki, or Pashto; of these Balochi is the most important. (On “bilateral bilingualism,” see Elfenbein 1990: 438.) Despite the fact that Balochi and Brahui are genetically unrelated (Brahui is Dravidian, Balochi is Indo-European Iranian), a linguistic symbiosis has developed between the two

AUTHOR’S NOTE: Abbreviations as in the chapter on Balochi, plus the following: abl., ablative; acc., accusative; Bal, Balochi; conj., conjunctive; dat., dative; *DED*, *Dravidian Etymological Dictionary*; Dr, Dravidian; inv., imperative; inst., instrumental; Lhd, Lahndā; n., nominative; loc., locative; pfx., prefix; pret., preterite; prohib., prohibitive; Si, Sindhi; tr, transitive.

languages at least since the 15th century (and likely much earlier, see Elfenbein 1982, 1987, 1990: 438). This has resulted in a profound “Balochification” of Brahui, most obviously in its phonology; but there are other, morphological, syntactical, and especially lexical influences as well (some of these are described in Elfenbein 1982).

Still the best study of Brahui matters as a whole (language, tribes and tribal life, geography) is to be found in Sir Denys Bray’s *The Brahui Language I (Grammar)* (1909), *II The Brahui Problem*, *III Brahui Etymological Dictionary* (1934). Elfenbein 1990 includes a full bibliography up to 1985. Traditionally, it has been held that the Brahuīs have lived in the highlands of Kalat since prehistoric times, having split off from other Dravidians before their arrival in India; but I have proposed a revision of this notion to one of a relatively late immigration from India, about the 9th century (Elfenbein 1987). According to this theory, they met and became mixed with the Baloch in their present habitat about a thousand years ago, and the pervasive influence of Balochi on Brahui dates from about this time.

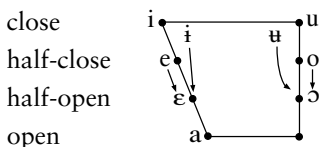
39.2. Dialects

Brahui is essentially without important dialectal variation. One phonological difference only distinguishes northern (Sarāwān) Brahui from Southern (Jahlawān) Brahui, with the dividing line running from east to west somewhat south of Kalat: this is the treatment of /h/. In the north it is pronounced as [h] in all positions, whereas in the south it is usually weakened to a glottal stop [ʔ], written ’, or Ø. Kalati Brahui, a sort of standard, keeps /h/, see § 39.3.3.1.

39.3. Phonology

Segmental phonemes (see Bray I, 23–24; cf. also Elfenbein 1990: 439).

39.3.1. Vowels



The same phonetic remarks hold as for the vowels of Balochi (§ 37.4.1). Non-phonemic variation is indicated by arrows. In addition to the variations

shown in the Balochi quadrilateral, Brahui has an additional allophone of *-a* [a] in final position, where it tends to [æ]—but this varies with speaker, and tends to be commoner in the north than in the south. In Bray's description of the vowels, an additional short /e/ appears, effectively disposed of in Emeaneu 1962: 7 as an allophone of either [e:] or [i:]. Hence the Brahui segmental vowel phonemes are:

Short	<i>a i u</i>
Long	<i>ā ī ū ē ō</i>

exactly as in Balochi. The long vowels are approximately twice as long as the short. Stressed *ī* and *ē*, *ō* and *ū* are kept apart in all Brahui speech, but unstressed /*ō*/ → [ɔ:], and unstressed /*ē*/ → [ɛ:]. Stress patterns have been taken over from Balochi: a long vowel or diphthong attracts stress, and in polysyllables stress is on the first long vowel; where all vowels are short, it is on the first syllable.

Minimal pairs for *ō* : *ū* and *ē* : *ī* as distinct phonemes are:

<i>sōb</i> : <i>sūb</i>	‘victory’ : ‘morning’
<i>kōrk</i> : <i>kūr</i>	‘stony ground’ : ‘soft wool’
<i>kōš</i> : <i>kūš</i>	‘camel thorn’ : ‘a girl’s game’
<i>ē(d)</i> : <i>ī</i>	‘that’ : ‘I’
<i>dē</i> : <i>dī</i>	‘day’ : ‘then’
<i>lēkō</i> : <i>līkō</i>	‘for the sake of’ : ‘a type of work-song’

whereas in final position, *-ū* and *-ō* appear to be free variants (see § 39.7.3) in 3 pl. verbal endings, otherwise we have e.g. the diminutive suffix *-ū* (< Persian) *gōsū* ‘calf’, *piššū* ‘kitten’, with no *-ō* variant; the prohib. 2 pl. inv. *-bō* ‘do not ...’ has no *-bū* variant, etc.

39.3.2. Diphthongs

Phonemically, there are only two diphthongs, as in Balochi: /ay, aw/ with slight allophonic variations (see Elfenbein 1982: 82) which as a rule do not occur in Balochi: /ay/ is mostly [a:i] varying to [æ:i]. Bray's *ei* is an allophone of /ay/, occurring but seldom: *bai*, pl. *baik* [bæi, bæik] ‘a grass’, *antay* [antæi] ‘why?’, *say* ‘3’ [sæi]; *sayl* [sæ:il, sæil] ‘spectacle’: it is significant that the last two words are loanwords from Ra Balochi, see § 37.4.2, where the same pronunciation occurs. The variants given by Bray *sei*, *seil* [sei, sæil] are the expected southern Ra variants.

The diphthong /aw/ [a:u, æ:u] tends to [o:], especially in the south, again influenced by neighboring Co Balochi dialects (see § 37.4.2). In the areas where *aw* is a monophthong, oppositions such as *maun* : *mōn* ‘black’ : ‘face’,

maur: *mōr* ‘a wild oil seed’ : ‘peacock’ are lost. Final *-aw* is often lengthened to [a:u], sometimes with some friction [a:ɰ]. Otherwise, as in Balochi, friction is minimal. This diphthong is kept apart from *āū/ō*.

39.3.3. *Segmental consonant phonemes*

See Emeneau 1937, Elfenbein 1982, 1983, 1990.

Table 39-1. *Brahui Consonants*

	Labials	Dentals	Palato-Alveolars	Retro-flexes	Velars	Glottals
Stops	p b	t d		ʈ ɖ	k ɡ	ʔ
Affricates			ʈʃ ɖʒ			
Fricatives	f				x ɣ	h
Spirants		s z	s ʒ			
Nasals	m	n		ɳ	(ŋ)	
Laterals		l ɭ				
Flaps		r		ɽ		
Semivowels	w	y				

The articulations are identical to Balochi (§ 37.4.4); only the following require comment.

39.3.3.1. *Glottals*

h Voiceless glottal fricative [h]—but with little friction; it is retained in all positions in the north, but is unstable in the south, where it tends to a voiceless glottal stop [ʔ] in initial and intervocalic position, and is lost before consonants and in final position:

North	South	
<i>hust</i>	<i>ʔust</i>	‘heart’
<i>himpa</i>	<i>ʔimpa</i>	‘don’t go!’
<i>sahī affaʔ</i>	<i>saʔī affaʔ</i>	‘I don’t know’
<i>kēhalk tammā</i>	<i>kēʔalk tammā</i>	‘he stumbled (and) fell’
<i>pōh</i>	<i>pō</i>	‘intelligence’
<i>śahd</i>	<i>śaʔd, śad</i>	‘honey’

ʔ glottal stop, a positional variant of [h] in the south, initially and intervocalically. It is often prefixed to vowel-initial words in the south.

There is a common perception that the retention of *h* represents a “better” pronunciation.

39.3.3.2. *Fricatives*

The fricatives *x*, *ɣ*, and *f*, which occur in Balochi only in recent loanwords (except in the EHB dialect, see §§ 37.3.6 and 37.4.4) occur in Brahui in native words in their own right, without allophonic variation; all of the following examples have Dravidian etymologies in *DED*:

<i>ōfk</i>	‘they’
<i>bārifing</i>	‘to make dry’
<i>xulis</i>	‘fear’
<i>xāxar</i>	‘fire’
<i>mux</i>	‘waist’
<i>ǰaxxing</i>	‘to pierce’
<i>ɣuḍḍū</i>	‘small’
<i>lummayāt</i>	‘mothers’
<i>pirɣpa</i>	‘do not break’
<i>kēraɣ</i>	‘bottom’
<i>kas(i)fiŋg</i>	‘to kill’

Initial *f*-, however, occurs only in loanwords from NP/Ar, and tends to *p*-.

39.3.3.3. *Velar nasal*

The sound [ŋ] occurs only before *k* and *g*; but the spelling *-ing* of the ending of the verbal noun means [iŋg]; *ŋ* becomes [m] before *p* and *b*: *harsingpa* = [harsimpa] ‘do not turn around’.

39.3.3.4. *Voiceless lateral*

The voiceless lateral *lh* [ɬ], articulated with some friction, does not occur in any Balochi dialect, nor in any neighboring language. Its appearance is limited to postvocalic position in a few words (all are listed below), and it is opposed to *l* in all except one or two cases:

<i>tēlh</i> : <i>tēl</i>	‘scorpion’ : ‘oil’
<i>mēlh</i> : <i>mēl</i>	‘sheep’ : ‘assembly’
<i>pālh</i> : <i>pāl</i>	‘milk’ : ‘omen’
<i>mōlh</i> : <i>mōl</i>	‘smoke’ : ‘muffling’
<i>malh</i> : <i>mal</i>	‘son’ : ‘wrestling’
<i>hīlh</i> : <i>hīl</i>	‘fly’ : ‘hope’
<i>sēlh</i> : <i>sēl</i>	‘winter’ : ‘spectacle’ (in areas where <i>ay</i> > <i>ē</i>)

Its origin is obscure; it seems to arise, along with *l*, from both Proto-Dravidian **l* and **ɭ*, in conditions as yet unknown; e.g., in *hilh* ‘fever’ (no minimal pair), *lh* can represent equally either PDr **l* or **ɭ*, whereas in *pilhing* ‘to squeeze’ (no minimal pair) *lh* represents PDr **l* (PDr **ɭ* ordinarily remains in Brahui). In inflexion *lh* is preserved:

<i>tē<u>lh</u>nā</i>	g. sg. [te:ɭna:] ‘of a scorpion’
<i>tē<u>lh</u>k</i>	n. pl.
<i>tē<u>lh</u>tā</i>	g. pl.

Noteworthy is the verbal stem *xalling*, with meanings (1) ‘to strike’, (2) ‘to steal (cattle)’, (3) ‘to uproot (plants)’, in which, according to *DED*, three different Dr roots have coalesced: (1) has a 2 sg. inv. *xalh* ‘strike!’ (cf. *xal* ‘stone’ and *xalh* ‘pain’ (< ‘stricken’), misprinted in Bray III, 179 as *khallh*²; the other forms of (1), as well as all forms of (2) and (3), show only *xal*-. Two other verb stems show this peculiarity: *tūling* (stem *tūl*-) ‘sit’, has inv. 2 sg. *tūlh*; *halling* (stem *hal*-) ‘seize’ has inv. 2 sg. *halh*.

39.3.3.5. Retroflexes

In contrast to the case in Balochi, where retroflexes do not occur in native words except as positional variants, in Brahui we have in native words (i.e. with *DED* etymologies):

ɭ [ɭ] in 1st person verbal endings *-uɭ*, *-aɭ*, *-ōɭ*; *iraɭ* ‘two’, *buɭ* ‘hillock’, *muɭɭux* ‘knotted’, etc.

ɖ [ɖ] in *piɖ* ‘belly’, *guɖɖū* ‘small’, etc.

ɾ [ɾ] (does not occur in initial position) in *bīɾing* ‘to milk’, *pēɾing* ‘to wrap round’, *dēɾū* ‘borrowing of milch animals’, *harsing* ‘to turn round (tr.)’, etc.

There is a tendency in EHB-influenced northern Brahui to develop *r* before *t*, *d*, *s*, *z* to *ɾt*, *ɾd*, *ɾs*, *ɾz*:

<i>xūɾ/ɾt</i>	‘tiny’
<i>har/ɾtoma</i>	‘both’
<i>mard/maɾd</i>	‘man’
<i>lurd/luɾd</i>	‘turbid’
<i>har/ɾsing</i>	‘to turn around’
<i>lar/ɾzing</i>	‘to tremble’
<i>dar/ɾzī</i>	‘tailor’
<i>har/ɾdē</i>	‘every day’
<i>par/ɾda</i>	‘curtain’

ɾ can alternate with *ɖ* in:

<i>bōṛ ~ bōḍ</i>	‘louse’
<i>pōṛ ~ pōḍ</i>	‘clod’
<i>gēṛ ~ gēḍ</i>	‘roofed enclosure for animals’
<i>dāṛān ~ dāḍān</i>	‘from this (abl.)’

This tendency is weaker in urban speech.

39.4. Aspiration

Confined to the stops *p t k* in the same conditions as in EHB (§§ 37.3.6, 37.4), it is heard mainly in the northeast, where it is a free variant and is non-phonemic:

<i>pōk ~ phōk ~</i>	
<i>phōkh</i>	‘wasted’
<i>tōba ~ thōba</i>	‘repentance’
<i>kūring ~ khūring</i>	‘to roll up’
<i>ṭikaṭ ~ ṭhikaṭh ~</i>	
<i>ṭhikhaṭh</i>	‘ticket’

The phenomenon is certainly to be ascribed to EHB influence. Also occasionally heard are the aspirates *čh, bh, dh, gh, jh* in loanwords from IA, mainly in initial position, in the south: there is free variation with the non-aspirated form:

<i>čhār ~ čār</i>	‘marsh’, cf. Si <i>chōra</i> ‘id.’
<i>jhil ~ jil</i>	‘a plant’, cf. Si <i>jhilu</i> ‘id.’
<i>dhak ~ dak</i>	‘a blow’, cf. Si <i>dhaku</i> ‘id.’
<i>dhōbī ~ dōbī</i>	‘washerman’, cf. Si <i>dhōbbī</i> ‘id.’
<i>gharī ~ garī</i>	‘clock’, cf. Si <i>gharī</i> ‘id.’
<i>ghōṛaw ~ gōṛaw</i>	‘troop of horse’, cf. Lhd <i>ghōṛaw</i> ‘id.’

Expressive aspiration is also heard; mainly in initial position:

<i>bhāz</i>	‘many’
<i>bhōp</i>	‘mattress’
<i>ghaṭ</i>	‘inaccessible’

39.4.1. Nasalization

There is a strong tendency to nasalize *ā* before *z*, particularly; but other vowels can also be nasalized before *z*:

<i>bāz</i>	‘hawk’
<i>bāzil</i>	‘upper arm’
<i>dāz</i>	‘dust’
<i>sirōz</i>	‘fiddle’

but all of these words can be loanwords from EHB.

čã is certainly a loanword from EHB. The numerals 11 to 16 are all borrowed from Balochi (EHB): *yãzda*, *dwãzda*, *sẽzda*, *čãrda*, *pãzda*, *šãzda*; other numerals (except 1, 2, 3) are also borrowed from Balochi, but the dialect of origin is not certain.

39.5. Epenthesis

39.5.1. Vowel epenthesis

Monosyllables with either a final vowel or consonant infix an epenthetic *-a-* before suffixed endings with an initial consonant, as in the following:

<i>bãh</i>	‘mouth’
<i>mãr</i>	‘son’

g. sg. *-nã*, conj. sg. *-tõ*; we have e.g. *bãh-a-nã*, *bãh-a-tõ*, *mãr-a-nã*, *mãr-a-tõ*. In the south, one hears *bãʔanã*, *bãʔatõ*. Consonant clusters can be opened, also, with *-a-*:

<i>gidisp-a-nã</i>	g. sg. ‘span’, beside <i>gidispnã</i>
<i>kapõt-a-tõ</i>	conj. sg. ‘pigeon’, beside <i>kapõttõ</i>

As in Balochi, vowels in close juncture (§ 37.4.3) have neither hiatus nor friction:

- a) monosyllables in long *ã* suffix case endings with an initial vowel:

dat./acc.	<i>-ẽ</i>
abl.	<i>-ãn</i>
inst.	<i>-aʔ</i>
loc. 1	<i>-aʔĩ</i>
loc. 2	<i>-ãĩ, -ã</i>

without hiatus: *ã - ã / ãã* with 2 + 2 morae; e.g., (Southern dialect) *bãẽ*, *bããn*, *bãaʔ*, *bããĩ*, *bãã*.

- b) Monosyllables ending in other vowels:

<i>dẽ</i>	‘day’
<i>xõ</i>	‘cooking pot’
<i>dũ</i>	‘hand’

suffix these endings without epenthesis.

- c) Monosyllables in a diphthong: *bai* ‘a grass’: *baiẽ*, *baiãn*, etc., have *-y-* [j] epenthesis: [*bajẽ*] [*bajãn*]; but: *bãĩ* ‘cover’: *bãĩẽ*, *bãĩãn* [*ba:i:je*], [*ba:i:ja:n*].

Bases in *-aw*: before a vowel ending, [-aʊ]; before a consonant ending, [-au-]; e.g., *kulaw* ‘message’, n. pl. [*kulawāk*], abl. sg. [*kulaunã*].

39.5.2. Consonant epenthesis

39.5.2.1. Epenthetic -y-

By far the most common element is -y- (not counting [-j-] and [-w-], § 39.5.1c). Bisyllables in -ā have a -y- epenthesis before case endings with an initial vowel:

<i>lumma</i>	‘mother’, n. pl. <i>lumma-y-āk</i> (* <i>lumma-āk</i>), etc.
<i>bāva</i>	‘father’, n. pl. <i>bāva-y-āk</i>
<i>parra</i>	‘wing’, n. pl. <i>parra-y-āk</i>
<i>urā</i>	‘house’, n. pl. <i>urā-y-āk</i>

Other endings: *lumma-y-ān*, *lumma-y-āi*, etc.; *urā-y-ān*, *urā-y-āi*, etc.

Adjectives in -ā have a -y- epenthesis before the definite attributive ending -ā (polysyllables)/-angā (monosyllables), e.g.:

<i>zēbā-y-ā masīr</i>	‘the pretty daughter’ (* <i>zēbā-ā</i>)
<i>ganda-y-ā mār</i>	‘the bad son’ (* <i>ganda-ā</i>)

(I have not found any monosyllabic adjectives in -ā with a suffix -y-angā.)

In four verbs, y also appears epenthetically before vowels:

<i>sa-y-ing</i> (stem <i>sa-</i>)	‘to pound’ (Bal <i>sā-</i> , <i>sāita</i> ; cf. NP <i>sāĩdan</i>)
<i>hō-y-ing</i> (stem <i>hō-</i>)	‘to weep’
<i>bē-y-ing</i> (stem <i>bē-</i>	
< pfx. <i>bē</i> ‘over’)	‘knead’ (‘do over’)
<i>šā-y-ing</i> (stem <i>šā-</i>)	‘to put in’

39.5.2.2. Epenthetic -r-

Epenthetic -r- is to be found in *ama-r-*: *ama* ‘how?’ before the copula verb and before suffixed -ī (§ 39.6.3). E.g., *ō ama-r ē* ‘how is he?’ *nā ama-r-ī ant ē* ‘how are you?’ (‘your howness what is it?’); *-bā-r-*: suffix *-bā* ‘like’; *ōnā mōn his-bā-r ē* ‘his face is like ashes’; *nēām-bā-r-ō sardaras* ‘a chief like you’ (with indefinite adjective suffix -ō).

-r- epenthesis is probable in the present indefinite negative conjugation *-pa-r* (negative conjugation infix *-pa-*) 1 sg. ending, leveled from forms like those of the probable future conjugation, where it is certain: 1 sg. *-pa-r-ōt*, 2 sg. *-pa-r-ōs*, etc. In the past conditional it is also clearly present: 1 sg. *-pa-r-ōs-ut*, etc., where -ōs- is the tense-sign.

39.5.2.3. Epenthetic -n-

The case suffixes *-tō*, *-ki*, *-iskā*, *-ēkā* suffix an epenthetic -n- before a following substantive verb, both affirmative and negative:

kanā mārātō-n uṭ 'I am with my son'
kantō-n aff 'it is not with me'
kanā ilumki-n ē 'it is for the sake of my brother'
kanā banningiskā-n
aḍ ka 'wait until I come!' ('wait for my coming')
ofk mastē ka-n ō 'they are as far away as the hills'

The adverbs: *dākā* 'hence', *ōkā* 'thence (mediate)', *ēkā* 'thence (remote)', *arākā* 'whence?', *dārēskā* 'up to here', *ōrēskā* 'up to there (mediate)', *ērēskā* 'up to there (remote)' all take a suffixed *-n* before a following word beginning with a vowel.

39.6. Consonant gemination

39.6.1. In monosyllables

As in Balochi, monosyllables with a short vowel can have an expressive gemination of the final consonant, for emphasis. Particularly affected are *-č*, *-x*, *-t*, *-l*. But gemination of a final consonant can also be phonemic, as the following minimal pairs show:

<i>čaṭ : čaṭṭ</i>	'ruined' : 'lazy'
<i>šir : širr</i>	'lentil' : 'evil spirit'
<i>gar : garr</i>	'gorge' : 'mange'
<i>kad : kadd</i>	'manure' : 'height'
<i>kas : kass</i>	'someone' : 'kinsman'
<i>lak : lakk</i>	'lakh (100,000)' : 'hill pass'
<i>šak : šakk</i>	'doubt' : 'peak'

but these are not common. There is a syllable balance rule (as in Balochi) which opposes long vowel + single consonant to short vowel + geminated consonant. E.g.:

<i>pākī : pakkī</i>	'cleanliness' : 'small palmful'
<i>pāra : parra</i>	'side' : 'feather'
<i>pūr(a) : purr</i>	'full' (loanword < IA) : 'full' (loanword < Bal)
<i>sāg : sagg</i>	'greens' : 'endurance'
<i>sāring : sarring</i>	'to keep' : 'to grow up'

There are a few exceptions to this rule:

<i>mačī : maččī</i>	'like a date palm' : 'fish'
<i>sat : sattī</i>	'bravery' : 'brave'
<i>tilī : tillī</i>	'splinter' : 'spleen'

The gemination of a final voiceless stop after a short vowel is not distinguished in pronunciation before a following vowel-initial word, from its

pronunciation in a close juncture sequence of two words the first of which has a final voiceless stop identical to the initial stop of the following word; i.e., there is no difference in pronunciation between -KK V- and -K K- (K = *p, t, k*). This circumstance has led Emeneau to a new (and to me convincing) explanation for the origin of the initial *k*- in oblique forms of the 1 sg. personal pronoun (n. sg. *ī*, obl. stem *kan-*, see Emeneau 1991).

39.6.2. Phonemic status

Gemination is phonemic. The following are illustrations:

<i>dukān</i> : <i>dukkān</i>	‘eye’ (abl. sg.) : ‘shop’
<i>gilik</i> : <i>gillik</i>	‘red ochre’ : ‘buttock’
<i>mana</i> : <i>manna</i>	‘some’ : ‘temporary summer shelter’
<i>maza</i> : <i>mazza</i>	‘big’ : ‘taste’
<i>piring</i> : <i>pirring</i>	‘to swell’ : ‘to twitch’
<i>rabī</i> : <i>rabbī</i>	‘spring crop’ : ‘a kind of date’
<i>katār</i> : <i>kattār</i>	‘line’ : ‘knife’
<i>xuling</i> : <i>xulling</i>	‘to fear’ : ‘to pierce’

39.6.3. Suffix -ī

Before the suffix -ī (forms adjectives from nouns, or nouns from adjectives, a loan morpheme from Balochi or Persian), a final consonant is always geminated:

<i>buḷ</i> : <i>buḷḷī</i>	‘plugged up’ : ‘a plug’
<i>čup</i> : <i>čuppī</i>	‘silent’ : ‘silence’
<i>čur</i> : <i>čurrī</i>	both ‘streamlet’
<i>kuḍī</i> : <i>kuḍḍī</i>	both ‘hut’ (<i>kuḍī</i> has its -ī < <i>kuḍḍī</i>)
<i>daḍ</i> : <i>daḍḍī</i>	‘strong’ : ‘strength’
<i>xwaš</i> : <i>xwaššī</i>	‘happy’ : ‘happiness’

There are countless oppositions of the sort

<i>čut</i> : <i>čuttī</i>	‘a boil’ : ‘escape!’ (2 sg. inv. < <i>čuttīng</i>)
---------------------------	---

involving all consonants, where the emphatic pronunciation of the first word eliminates the opposition.

39.6.4. Preceding other suffixes

Final -*t*, -*ḍ*, -*č*, -*p* are geminated when suffixed by an ending with an initial vowel or diphthong:

<i>laṭ</i> ‘stick’	dat. sg. <i>laṭṭē</i> , conj. sg. <i>laṭṭatō</i>
<i>puč</i> ‘clothes’	abl. sg. <i>puččān</i> , loc. 1 sg. <i>puččatī</i>

The n. pl. suffix is *-k* or *-āk* depending on the final base vowel or consonant, with the following exceptions due to assimilation:

xad ‘nipple’ n. pl. *xakk* (*kk* < *dk*)

nat ‘foot’ n. pl. *nakk* (*kk* < *tk*)

(other bases in *-d* or *-t* form n. pl. in *-āk*); all bases in *-r* with a short vowel in the last syllable assimilate *rk* > *kk*:

šahr ‘town’ n. pl. *šahkk*

kur ‘flock’ n. pl. *kukk*

ditar ‘blood’ n. pl. *ditakk*

xāxar ‘fire’ n. pl. *xāxakk*, etc.

but if the last syllable has a long vowel, the *-r* is simply dropped:

mār ‘son’ n. pl. *māk*

ḍayār ‘soil’ n. pl. *ḍayāk*

layōr ‘coward’ n. pl. *layōk*

angūr ‘grape’ n. pl. *angūk*

Other plural case endings are suffixed to the base + *tē-* (except g. pl. *-tā*). When the base ends in *-t*, *-d*, or *-r* the plural base is in *-tt-* (by assimilation from *dt*, or *rt*, respectively): *xad* has a pl. oblique base *xatt-*; *nat* has *natt-*, *kur* has *kutt-*, and *xāxar* has *xāxatt-* (further details, see Bray I, §§ 25–38).

39.6.5. *-n/r-* Stems

For the *-nn-* in the common verbs *manning* ‘become’, *banning* ‘come’, *kanning* ‘do’, *danning* ‘take away’, *anning* ‘be’, *hunning* ‘look’, see Emeneau 1962, ch. 3, “The *n/r*-stem problem.” On the other hand, assimilation of *rp* > *pp* is clear in the present negative verbal stems *kappa-* (*kanning*), *dappa-* (*danning*), *huppa-* (*hunning*); the other verbs have a different formation in *mafa-* (*manning*), *bafa-* (*banning*) and *affa-* (*anning*).

39.7. Loss of finals

39.7.1. Vowels

Loss of final vowels does not seem to occur. The loc. 2 ending *-ā* has been explained by Bray as a shortening (i.e., a loss of *i*) in the ending *-āi*; but much more likely is that the ending is simply the Balochi *-ā* case ending. At the very least an influence from Balochi is certain.

39.7.2. Loss of final consonants

In the n. sg. of the demonstrative pronouns *dād* ‘this’, *ēd* ‘that’, *ōd* ‘this, that’, the final *-d* is lost when the next word begins with a consonant. The *-d* is

always kept before the copula.

39.7.3. *-ur* > *-ū*

The 3 pl. past verbal ending *-ur* tends to drop *-r* and lengthen *u* to *ū*, which varies freely with *ō* (§ 3.1); e.g., *assur*, *assō* ‘they were’, etc. The perfect 3 pl. ending *-(u)nō* < **-(u)nur*: e.g., *ōfk kungunō* ‘they have eaten’ (**kungunur*).

39.8. Present adverbial participle

Bray’s “present adverbial participle” in *-isa* (I, § 318), e.g., *bin-isa* (1) ‘hearing’, (2) ‘picking up’ can take a suffixed *-ū/-ō* (§ 3.1) which gives a past active force (see Elfenbein 1983: 106); e.g., *ō kasarat činjik binisaū/ō hinaka* ‘he went along the road picking up sticks’ (i.e., ‘picked and he went’); *kana urā hōyaū/ō pare* ‘my wife wept and said’ (< *hōya-ū/ō* 3 sg. pret. ‘she wept and ...’).

This construction is in essence the same as that in EHB Balochi, in which when two consecutive past-participle verbs (transitive or intransitive) occur, the first has an *-o* “suffix” and they describe two consecutive or contemporary past acts: EHB *maĩ jan grēθ-ō gušt* ‘my wife wept and said’, (i.e., ‘my wife, having wept, said’). The *-ō* (as was already recognised by Bray) is none other than *ō* ‘and’, and thus is likely to be a borrowed construction in Brahui from EHB.

39.9. Metathesis

A few common words are often metathesised as in Balochi; the direction of the influence is not certain. We have, e.g., *matbal* < *matlab* ‘meaning’, *rusxat* < *ruxsat* ‘permission’, *kulf* < *kuf* ‘key’, *urz* < *uzr* ‘excuse’, and a few others.

39.10. Rhyme words

There is an expressive use of rhyme to form plurals in e.g. *puč-muč* ‘clothes’, *zāifa-māifa* ‘womenfolk’, apparently only with *m-*.

39.11. Stress

Stress is fairly strong but is non-phonemic. A long vowel attracts stress; polysyllables of short vowels are stressed on the first syllable. Polysyllables with more than one long vowel are stressed on the first one.

39.12. Lexicon

Approximate word counts, using Bray, *DED*¹, *DEDS*, and Elfenbein 1983, show that of the Brahui lexicon

- 10% is of Dravidian origin
- 20% is of IA origin
- 20% is of Balochi origin
- 30% is of NP/Ar origin
- 20% is of unknown origin

Bibliography

General bibliographies of Brahui studies are to be found in *Encyclopaedia Iranica* s.v. "Brahui," and in

Andronov, M. 1980. *The Brahui Language*. Moscow: NAUKA.

The following have valuable details on Brahui phonology:

Bray, Sir Denis. 1909–1934. *The Brahui Language*. I (Calcutta, 1909); II, *The Brahui Problem*; III, *Brahui Etymological Dictionary* (Delhi, 1934). Reprinted Quetta, 1972–1978.

Burrow, T., and Murray B. Emeneau. 1984. *Dravidian Etymological Dictionary*, 2nd ed. Oxford: Clarendon.

———. 1968. *Dravidian Etymological Dictionary*: Supplement [to 1st ed., 1961]. Oxford: Clarendon.

Elfenbein, Josef. 1982. "Notes on the Balochi–Brahui Linguistic Commensality," *Transactions of the Philological Society* 77–98.

———. 1983. "The Brahui Problem Again," *Indo-Iranian Journal* 25: 103–32.

———. 1987. "A Periplus of the Brahui Problem," *Studia Iranica* 16: 215–33.

———. 1990. "Brahui." *Encyclopaedia Iranica* 3: 433–43.

Emeneau, Murray B. 1937. "Phonetic Observations on the Brahui Language." *Bulletin of the School of Oriental (and African) Studies* 981–83.

———. 1959. *India and Historical Grammar*, pp. 40–71. Annamalainagar: Annamalai University.

———. 1962. *Brahui and Dravidian Comparative Grammar*. Berkeley & Los Angeles: University of California Publications in Linguistics.

———. 1963. *A Sketch of Dravidian Comparative Phonology*. Berkeley & Los Angeles: University of California Publications in Linguistics.

- . 1970. *Dravidian Comparative Phonology*. Annamalainagar: Annamalai University.
- . 1991. 'Brahui Personal Pronouns,' in *Studies in Dravidian and General Linguistics*, pp. 1–12. Hyderabad: Osmania University, Centre of Advanced Study in Linguistics.

This page intentionally left blank.

4 Nilo-Saharan Languages

This page intentionally left blank.

Nilo-Saharan Phonology

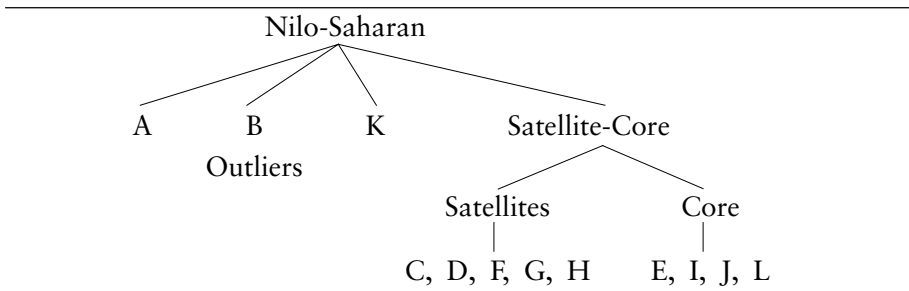
M. Lionel Bender

Southern Illinois University–Carbondale

40.1. Background

Nilo-Saharan is a phylum of languages located along the African Sahelian belt from Mali to Ethiopia.¹ Knowledge of the existence of such a phylum, its name, and the major outlines of its structure originates with the work of Joseph Greenberg, summed up in his 1963 publication. Omitting a large literature, I refer here mainly to my own 1989a and 1991 articles to carry the work forward to my present starting point. These articles embody a survey and analysis of morphological features (1989a: pronominals; 1991: other categories) of Nilo-Saharan (henceforth N-S) resulting in the genetic sub-classification shown in Table 40-1.

Table 40-1. Proposed Genetic Structure Of Nilo-Saharan



Abbreviations (names of the sub-families following Greenberg's original largely west-to-east order with some modifications arising from later work):

A. Songay	E. East Sudanic	I. Koman
B. Saharan	F. Central Sudanic	J. Gumuz
C. Maba	G. Berta	K. Kuliak
D. For	H. Kunama	L. Kado

1. AUTHOR'S NOTE: Most of the time and facilities to carry out this research were funded by the National Science Foundation under Grant No. DBS: 9021232: "Nilo-Saharan Reconstructions" in cooperation with Southern Illinois University at Carbondale.

NOTES: Maba is not to be confused with Mabaan, a Nilotic language (of Family E). For = Fur (choice of spelling). Greenberg included Families E, F, G, H in a larger family he called Chari-Nile. Later work (Goodman 1970 and Bender 1976a) did not support this grouping. I once considered Koman and Gumuz to be coordinates in a family, I-J Komuz, but now I consider them to be distinct enough to merit separate treatment. Kuliak is the original “Teuso,” formerly considered to be a branch of East Sudanic. “Kado” is the name suggested by the late Roland Stevenson for the “Kadugli-Krongo” or “Tumtum” family which Greenberg (1963) included in his Kordofanian family within another phylum (Niger-Kordofanian), although he expressed reservation about this detail (see 1963: 149).

The three outliers do not constitute a genetic group; they are coordinate with each other and with Satellite-Core (henceforth S-C). Details of the structure of the Satellites and S-C branches are unclear: structures proposed in my previous publications are undergoing revision as a result of current work, including this paper. Core families may be considered coordinates to each other. See Bender 1989a, 1991 for the morphological isoglosses which establish Core and S-C. The Core is the innovative group within N-S (e.g. in the areas of gender markers, inclusive/exclusive distinctions in pronouns, and nominal plurals). Other innovations separate S-C from the outliers. The three outliers are on the geographical fringes of N-S territory and, typical of such cases, they retain archaic N-S features.²

Accepting mutual intelligibility as the criterion of language identity, the Nilo-Saharan languages number at least 120. The exact numbers within the large Central Sudanic family and the large Nilotic sub-family of East Sudanic are problematical. The most likely N-S languages to be recognized by the non-Africanist linguist are probably Songay of Mali, Kanuri of northeast Nigeria, Nubian (the Egyptian varieties), and several Nilotic varieties (Masai, Nuer, Dinka, Luo).

These are also among the demographic leaders. According to my forthcoming survey, the total number of (first-language) speakers of Nilo-Saharan languages (in ca. 1988) is about 27,000,000, less than that of either Swahili

2. In Bender 1989a and 1991 I considered Families A, B, H, K to be “outliers,” but the current lexical work is making a case for H as part of the S-C grouping. This would lessen the problem of H (Kunama) having inclusive/exclusive distinctions in pronouns, a supposed innovation of the Core group. (However, the distinction is also found in K Kuliak—perhaps independent innovations; see Bender 1989a for discussion.) Regarding the structure of S-C, details are unclear and I prefer for the present to adopt a “neutral” scheme as in Table 40-1.

or Hausa alone (out of an African total of over 600,000,000). Of these, half (about 13,500,000) are Nilotic-speakers. The leading individual languages are Kanuri and Luo (ca. 3,500,000 each), Songay and Acoli (ca. 2,000,000 each), Kalenjin, Dinka, and Teso (ca. 1,500,000 each), and Nuer (ca. 1,000,000). All these are Nilotic except Kanuri and Songay.

The accompanying map shows the distribution of the Nilo-Saharan languages by family.³ §40.5 is a list of all known major varieties of Nilo-Saharan.

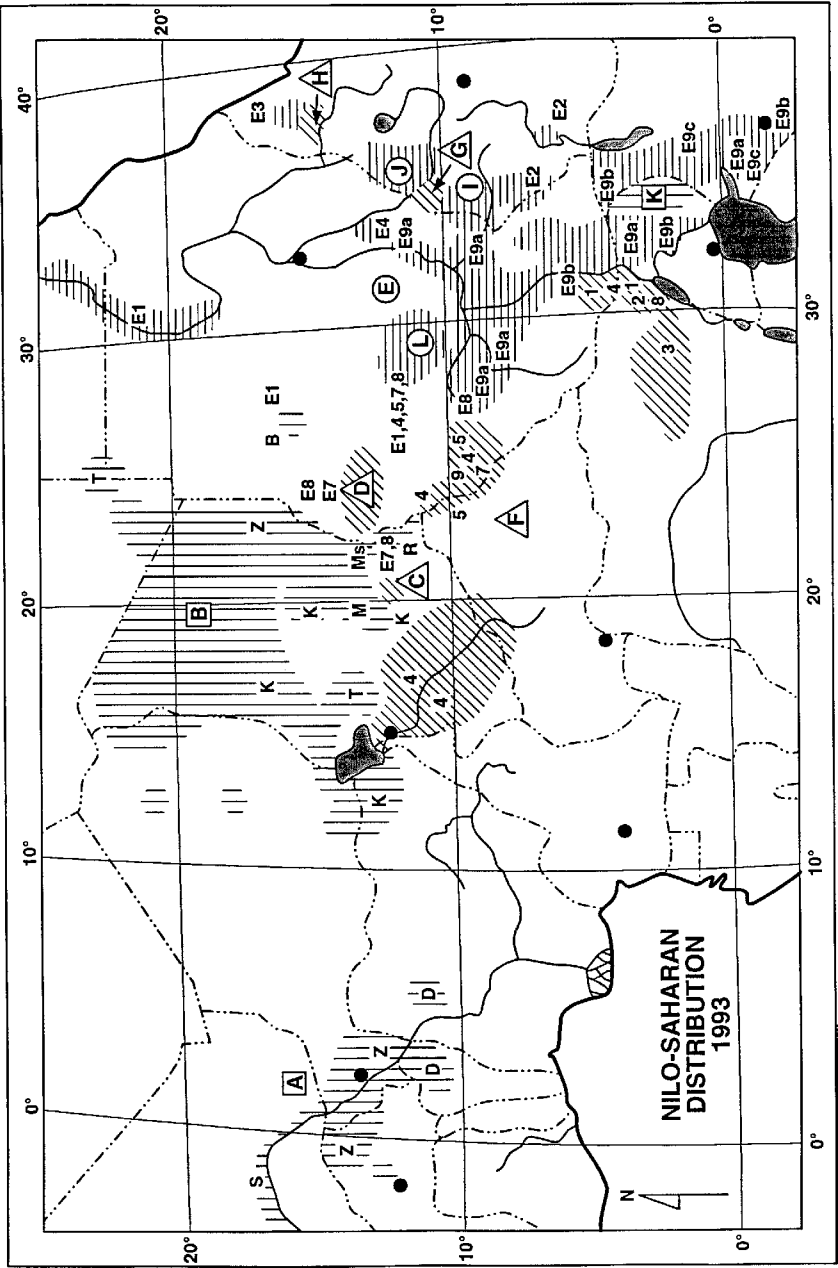
40.2. Introduction

The current work is based on collection and analysis of lexicon in about 115 languages, according to availability and completeness of data sets, representing quite well the range of variation in N-S. As it happens, the collection divides into four nearly equal groups: Nilotic (a sub-family of E, about 28 languages), E (East Sudanic) other than Nilotic, ca. 35 languages, F (Central Sudanic, ca. 24 languages), and all other N-S other than the preceding (ca. 28 languages). The lexicon used is a list of about 600 items arrived at by long experience as to what is available and useful. Each of the four sub-collections fills about 500 sheets, thus about 2000 sheets in all.

The lexicon is divided into 50 semantic sets which are derived from long experience and constantly subjected to modification. Examples of these sets are: 1 and 1a: quantity, size and distance respectively; 2, 2a, 2b: earth-like, locus, mountain-stone respectively; 3: bark and skin; 4: internal organs; 5: flying animals. Comparisons with high semantic plausibility are carried out within sub-sets and sets. “Crossovers” between or among semantic sets are routinely encountered, and taken into account, e.g. ‘fat’ (adj.) of set 1a with ‘fat’ (n.) of set 17: animals, flesh; “return” of set 14: motion towards with ‘give (back)’ of set 25: transfer.

From this analysis, a set of about 424 proposed isoglosses was identified (now reduced to about 208). These are at three levels: III: Nilo-Saharan, II: Satellite-Core, I: Core. The reconstruction work will be presented in Bender forthcoming. For other approaches, see Blench 1992, Ehret ms. and 1989. The proposed proto-segments, which will be referred to occasionally, are found in § 40.7.

3. The map is to be used for locating families. It is from an older classification scheme. Shading and labeling is somewhat out of line with the classification proposed herein.



Map 3. Nilo-Saharan Locations

NILO-SAHARAN DISTRIBUTION: LEGEND

Locations are approximate and not definitive. On the map, Ek and En groups are shown as E1, E2, etc.; Fp and Fc groups are shown by 1, 2, etc. Central locations of E = Ek + En and F = Fp + Fc are shown by  and  respectively.

OUTLIERS



Family
Songay 

D Dendi
S Songay
Z Zerma

Family
Saharan 

B Berti
K Kanuri
T Teda-Tubu
Z Zagawa

Family
Kuliak 
Ik
Nyangi
So (Tepes)

Rivers 

National boundaries 

Capital cities 

Lakes 

CORE GROUP



Family  ?

E1 Nubian
E3 Nera (Barea)
E5 Nylma ?
E7 Tama

Family 

E2 Surmic
E4 Jebel
E6 Temein ?
E8 Daju
E9 Nilotic
E9a (North)
E9b (Central)
E9c (South)

Family
Koman 

Family
Gumuz 

Family
Kado 
Kadugli-
Krongo-
Tulishi

SATELLITES



Family
Maban 

K Kibet
M Bora Mabang
Ms Mesalit
R Runga

Family
For 

For-Biltine

Family
Peripheral 
Central Sudanic

F1 Moru, etc.
F2 Mamvu, etc.
F3 Mangbetu
F7 Kresh-F9 Aja
F8 Baledha

Family
Core 
Central Sudanic

F4 Bagirmi, Sara
Yulu, Binga
Fongoro, Kara
Shemya (Sinyar)
Fer
F5 Bongo
F6 Modo, Baka

Family
Berta 

Family
Kunama-Ilit 

Map 3. Nilo-Saharan Locations

40.3. Common segments of Nilo-Saharan

The present paper is concerned with the synchronic state, i.e. the contemporary languages. From the published literature and some notes of my own and others, I prepared phonological charts of the segmental phonemes of the best-described languages with notes on major processes affecting these (mainly phonetic realizations; notes on suprasegmentals are also included). These will be condensed here in § 40.6. Note that for the large E and F families I list my own reconstructed proto-phonemes rather than go into the many languages in detail. Families E and F will be dealt with in Bender forthcoming (and for F, see also Bender 1992). Recall that E includes the large Nilotic sub-family, for which substantial reconstruction work has been done by other scholars.

Why is this project limited to segmental phonemes? I attempted from time to time to find generalizations involving suprasegmentals (length, phonation type, and especially tone) but I found the data so uneven both in terms of inclusion and reliability of suprasegmental marking that I gave it up except for a few notes on length (mainly of vowels) and phonation (e.g. aspirates, ejectives, implosives), as will be seen below.

The heart of this paper lies in § 40.6, to which I now turn in detail, following the major categories of segments: labials, dentals and alveolars, palatals, velars and post-velars, vowels. Experience shows that only two main syntagmatic positions need be considered: morpheme-initial (generally pre-vocalic) and post-vocalic (for convenience, sometimes referred to as medial).

40.3.1. *Labials*

The most interesting phenomenon in N-S labial phonemes is the frequent replacement of *p* by *f*. For this, see § 40.4.1.

Phonemic *b*, *m*, *w* are universal in contemporary languages. *b* and *w* are weaker post-vocalically than initially. In Soo and Nyangi of K (Kuliak) and Krongo of L (Kado) *b* is usually replaced by *β*, and in I: Twampa, *b* ~ *β*. Also in I, Komo, Kwama, and Opo *b* is initial only.

The glottalics *p'* and *β* are found mainly in S-C. Unlike Ehret, I could not reconstruct *p'*: it appears in the sub-group Twampa-Komo-Kwama of Koman, in alternation with *p* in Berta, and sporadically in several other groups with no clear distribution. *β* is a phoneme in F Central Sudanic, part of E East Sudanic, part of I Koman, J Gumuz, and L Kado, but not in C Maba (? see Nougayrol 1989: 19), D For, G Berta, or H Kunama. However, *β* is also strong in K Kuliak, perhaps an independent development.

The prenasal *mb* can perhaps be reconstructed in S-C: it is prominent in C, F, G, H, J, and L; less so in E and I. It also appears sporadically elsewhere, e.g., Kanuri.

The overall N-S pattern for labials thus appears to be: /p, b, f, m, w/ with a widespread later merger of *p* and *f* or replacement of *p* by *f*. The glottalics *p'* and *b* as well as prenasal *mb* are S-C innovations.

40.3.2. *Dentals and alveolars*

Dental-alveolar contrasts are cited for some languages, but the patterns are unclear in general so I do not attempt reconstructing the distinction. Universal here are /t, d, l, r, n, s/. Medial -*d*- and -*s*- are rare and initial *n*- and *r*- are very doubtful. The voiced fricative *z* occurs rarely but rather widespread among the families. Parallel to the labial case, implosive *d'* occurs for *d* in part of Kuliak (Soo and Nyangi). *θ* occurs in place of *s*, *z* in B: Tubu. *t* ~ *θ* dialectally in Berta. Occasionally, *s* ~ *z* in C: Bora Mabang. Parallel to the labials, *s* ~ *s'* in Berta. There is dialectal variation *s* ~ *h* in Kunama.

As with labials, glottalics /t', s', d'/ are perhaps reconstructable for S-C along with prenasals *nd* and *nt* (*s'* may very well be phonetic [ts']). Other phonemes of S-C families include aspirates (Koman), affricates (Kuliak), palatalized stops (Gumuz, Kado), retroflexes (Maba, Central Sudanic, Kado, perhaps Gumuz). There are also several varieties of *l* and *r* (reconstructable to different sources in *N-S). See § 40.6 and notes to it for details and also § 40.7 for proposed proto-phonemes.

The overall N-S system is /d, t, s, zʔ, n, l, r/ with later developments of /t', dʔ, s', nd, nt/ in S-C.

40.3.3. *Palatals*

All palatal phonemes are rare in N-S languages. Only *y* is universal, with *ɲ* nearly so: *ɲ* is not phonemic in B: Kanuri, occurring only before -*i* as a variant of *s*, and is absent in I: Komo, Opo, rare in Twampa and Kwama. Like *z*, all of *c*, *j*, (stops or affricates) and *ʃ* are widespread but rare; *ʒ* is very rare (reported in Songay, Central Sudanic, and Gumuz). Both *c* and *ʃ* are initial only. Even more sporadic are *c'* (in I, J, K), *j'* (implosive; in I: Twampa? and Gule and also in K?), and *ɲj* (C, H, J, L). Kuliak has palatal stops but no palatal fricatives.

Thus it seems that the overall system is /ɲ, y, c, j, ʃ/ and *ɲj* in S-C. With reservation, *c'* and *j'* may be ascribed to S-C. (*j'* is implosive and sometimes retroflex.)

40.3.4. *Velars and post-velars*

The universal set is *g, k, ŋ*. Medial *-g-* is lacking. The velar nasal often appears as *ŋg* and it is difficult to separate /*ŋg*/ from [ŋg] resulting from assimilation. Ejective *k'* is found in families G, I, J, K (in G, *k* ~ *k'*), and sporadically in E. The corresponding implosive *g'* is rare and limited to I, J, K. Other sporadic phonemes (e.g. labiovelars in some varieties of Songay) are seen in § 40.6.

Both *h* (initial only) and *ʔ* are widespread and one or both are found in every family except *East Sudanic (*Central Sudanic has marginal *ʔ*). Within East Sudanic, I found widespread instances of *h* in Ek and *ʔ* in En. Of course there are problems in asserting phonemic status to both these phones since they are strongly subject to alternations such as *ʔ, h* ~ zero, *ʔ* ~ *k'*.

The overall system then is /*k, g, ŋ, h, ʔ*/ with *k'* (and perhaps *g'*) being innovations of S-C (plus Kuliak).

40.3.5. *Vowels*

As has often been noted, vowels are subject to even more alternations than consonants and are more difficult to reconstruct. For Nilo-Saharan, the “usual five vowels” *i, e, a, o, u* are universal (except for A [Zerma] and L [Krongo]). Of these, none appears initially in roots except for a few instances of *a-* and even fewer of *o-*. All vowels appear frequently in prefixes.

There is much variation in the families as to quality, length, tension (or ATR), and gliding. Within Songay alone, Nicolai 1981: 38–45 identifies five vowel systems (and also 16 consonantal systems). Distinctive vowel length is reported for A, C (Aiki), D, G, H, I (Komo and Kwama), J, and L. Thus it may well be Proto-N-S. Distinctions in tension/ATR are clearest in C, K, L, and of course in the famous system of Nilotic (of family E). Diphthongs are reported most clearly for part of A, C, D, *E, H, I, J. The above distributions may best be interpreted as independent.

To summarize, the overall system is /*i, ii, e, ee, a, aa, o, oo, u, uu*/ with other developments not clearly distributed.

40.3.6. *The overall system*

Of course the above survey is very superficial—for more details, see § 40.6 and its notes, and for even more, go to the original sources (Tucker and Bryan 1966 [henceforth T&B], with most of its bibliography being found in Tucker and Bryan 1956, is an excellent starting point). The overall synchronic pattern is as in Table 40-2.

Table 40-2. Nilo-Saharan Segments Overall

Consonants					
	Labial	Dental/ Alveolar	Palatal	Velar	Post-velar
Stops					
+vce	b	d	j	g	
−vce	p	t	c	k	ʔ
Fricatives					
+vce		z			
−vce		s	ʃ		h
Nasals	m	n	ɲ	ŋ	
Sonorants		r, l			
Glides	w		y		
Vowels					
	Front	Central	Back		
High	i, ii		u, uu		
Mid	e, ee		o, oo		
Low		a, aa			
Innovated in Satellite-Core					
	ɓ	ɗ	jʼ	ɡʼ	
	pʼ	tʼ, sʼ	cʼ	kʼ	
	mb	nd, nt	ɲj	ŋɡ	

Note. The “overall pattern” is a fictitious entity: no specific language or family fits it exactly (unless by serendipity!) and it is not equivalent to the reconstructed proto-phoneme chart (for which see § 40.7). The pattern is intended to represent the “typical” N-S language in terms of frequency and possible historical saliency of phonemes.

This pattern is a very “bland” one: in the N-S overall pattern, there are no ejectives, implosives, prenasals, labiovelars, aspirates, etc. The postulated series of implosives/ejectives/prenasals in S-C may represent the results of contacts: ejectives and implosives from Afrasian, but the source of prenasals is unclear.

Refer to the map and consider the distributions:

Ejectives strongest in: G Berta, I Koman, J Gumuz, K Kuliak.

Implosives in: *F, I, J, K (also in E2 Surmic, E8 Daju).

Prenasals in: B (Kanuri *mb* and *ŋg* only), C, *F, G, H, J, L.

Except for Kuliak, ejectives are found in Ethiopia–Sudan border languages, an area of strong contact between Nilo-Saharan and Afrasian, notably Amharic (Semitic) and Oromo (Cushitic), both of which have ejectives. In ancient times, both of these were presumably not yet in the area, so the contact may be recent (within about the last 500 years). In earlier times, Agaw of Cushitic and Kefoid languages of Omotic may have been in the area: neither of these has implosives; Kefoid has ejectives.

The distribution of implosives is more problematical. I and J are in contact with Oromo. As just mentioned, earlier Kefoid and Agaw presence would not explain anything, since neither has implosives. F and K are far from the area and there is no plausible scenario of Cushitic–Central Sudanic contacts, though Kuliak was very likely in contact with Cushitic languages.

Finally, the distribution of prenasals is widespread and in no discernible pattern: it is most likely a case of independent development, given that nasal assimilations may easily give rise to prenasals in almost any language.

The situation of Kuliak is puzzling. Although Kuliak seems to belong with Songay and Saharan as isolated relics of old N-S, it shows similarity to Core N-S in having ejectives and implosives. This may be ascribed to ancient contacts with East or South Cushitic, but why then does Kuliak not partake in the “*f* for *p*” phenomenon? Heine (1976: 67–72) discusses Kuliak history and external relationships but leaves the question largely open, aside from a cautious suggestion of possible genetic relationship to Kalenjin (of Southern Nilotic). He shows that reconstructable lexicon suggests a quite old cattle/agricultural culture although some of the present-day peoples were encountered as hunter/gatherers. Fleming (1983: 426) takes a cautious position that Kuliak is East Sudanic. Ehret 1981 and 1989 considers Kuliak to be part of his very different East Sudanic.

40.4. Three brief topics in Nilo-Saharan phonology

40.4.1. The “*f* for *p*” Phenomenon

In a classic article, Ferguson (1976) presented “*f* for *p*” as a characteristic of the “Ethiopian Language Area,” originating in Afrasian: usually phonemic *p* is absent in the area languages, its place in the pattern being taken by *f*.

(He mentions that *v* is also missing, but I will not pursue this here.) The phenomenon is found in Semitic (including Arabic and Ethio-Semitic) and Cushitic (except Agaw), but is weak in Omotic (most Omotic languages have both *p* and *f*; see Bender 1989b: 136). According to Newman and Ma 1977: 9–10, Proto-Chadic had both *p* and *f* and the present loss of the *p*–*f* contrast in many present-day languages is due to (often independent) mergers. They do not mention Arabic contact as a cause or contributing factor.

It is worth examining the *p* for *f* distribution in Nilo-Saharan in some detail.

A Songay: One of the clearest cases of “*f* for *p*”: *p* is marginal, *f* predominates in the 16 various systems outlined by Nicolai 1981 (see my sample in § 40.6). Nicolai’s “structure fixe” (1981: 47) has *b* and *f*, no *p*.

B Saharan: The main varieties have *p* and *f*, but Hutchison (1981: 18) reports that in Kanuri *p* is an allophone of *b* or *f*.

C: In Mesalit, *p* (initial only) is a rare allophone of *f*, which latter itself is said to be rare (Edgar 1989: 9). For Aiki, Nougayrol (1989: 19) makes one of the clearest cases for *p* ~ *f*: that /*p*/ is realized as [p] or [f], the latter at first in Arabic loans and now expanding (paraphrased). Bora Mabang seems to have *p* but no *f* (Edgar 1989: 17, but T&B 194 gives *b* and no *p* or *f*).

D: For is another clear case: Jakobi (1990: 18) says *f* (allophones [f, ϕ , p]) replaces *p* in the stop series.

E: Briefly, *p* (initial only) occurs in all En languages. In Ek (E1 Nubian, E3 Nera, E5 Nyima, E7 Tama), *f* replaces *p*. This can be seen as reflecting the geography of Arabic (and perhaps other Afrasian) influence, though under this assumption, it is strange that E6 (Temein) and E8 (Daju) have been resistant compared to neighbors E5 and E7. Certainly E2 (Surmic) and E9 (Nilotic) have been most distant from Arabic influence and are accordingly least affected. Within E4, a gradient from both *p* and *f* to *f* for *p* is found.⁴

F Central Sudanic: The situation is summed up in Bender (1992: 17). The “peripherals” (“Moru-Madi, Mangbutu, Mangbetu, Kresh, Lendu,” using the names I used therein) are rather scattered on the edges of Arabic influence and have both *p* and *f*. But *p* is very rare in the Core Group, occurring only in Yulu and Bongoid, lacking in Bagirmi, Sara, Fongoro, Shemya and

4. Recall that “Ingessana,” a well-known name for the Gaam people, is supposed to mean ‘the thankless ones’ for their resistance to Islam (and Arabization). Of the three other Jebel varieties, it is my impression that Aka has more Arabic loanwords than the others, but a quick glance at the lexica seems to show Molo with *p* for *f*, Kelo with *p* only, and Aka and Gaam with both *p* and *f*! (See Bender 1994 for a followup).

the rest (for updated lists of Fp and Fc, see § 40.5.4). These are geographically nearer Arabic influence (see *ibid.* 17 for more details).⁵

G: Berta is an interesting case. The Berta are heavily Arabized, but their language retains *p* (though interestingly in alternation with *pʰ*) and also has *f* (my notes). T&B (348) gives two systems based on older material, one with both *p* and *f*, one with *f* only. In the description of Triulzi et al. 1976, *p* was replaced by *pʰ*, *k* by *kʰ*, and *t* by *θ*: in other words, there is no voiceless plain stop series. Obviously, Berta needs more investigation, since it seems to have much dialectal variation and the basic “dia-system” is not clear.

H: Kunama also shows the phenomenon. The Kunama today include Islamic, Christian, and “traditional” believers. Their present-day location, separated from other N-S (except for neighboring E3 Nera) in Eritrea, may be a result of flight from ancient Arab slave-raiding (see Bender and Alexander forthcoming).

I Koman: I reconstructed *p*, *pʰ*(?), *pʰ*, and *β*, but no *f* for Proto-Koman in my 1983a article. Present-day languages have various combinations, but only Gule (the presumably now extinct language of a thoroughly Arabized people) has *f* but no *p*.

J Gumuz: There are many clan varieties of Gumuz. For the best known, Uzar (1989: 348–50) documents a full series of labials: *p*, *pʰ*(rare), *b*, *β* (very rare), and *f*. For Sai I recorded all but *pʰ* in my 1979 description.

K Kuliak: Aside from the universal *b*, *m*, *w*, Kuliak has *p* and *β* throughout and *f* only in Ik.

L Kado: According to T&B (300), *p* but not *f* occurs in Kado as a family. Reh (1985: 11) lists both *p* and *f* but states (17) that *p* and *pp* are very rare (and incidentally, that *ff* is the only geminate consonant to appear initially).

To summarize: *p* for *f* is strong in: A, D, Ek, Fc, mixed in B, C, I, L, weak in En, Fp, G, J, K.

As usual in work of this kind, the picture is not one of a clearcut distribution, but one indicating a tendency. If *p* for *f* results from Afrasian contact (Arabic in all cases except for the extreme east, where Cushitic may play a role), then the first group makes sense. But in the last group, one wonders why Berta and Gumuz have not been more affected, especially Berta, since they have been so heavily Arabized compared to the Gumuz. All those in the mixed group (Koman less so) are in strongly Arabized areas and might easily

5. On the map, Central Sudanic languages are shown by bare numerals 1, 2, etc., whereas East Sudanic languages are shown as E1, E2, etc. The case for the peripheral nature of Fp (as regards Arabization) is not clear from the map alone and must be looked into in more detail to be convincing.

have been more affected. But obviously the mere fact of contact is not a *sufficient* explanation: other factors of history, linguistic structure, and culture surely play a part.

The mechanism of change of the sort at hand is multilingualism, especially Arabic–X bilingualism. The Sudan linguistic survey of the 1970s showed that Arabic is displacing other languages in Sudan at a rapid rate roughly from north to south. Among Nuba Hills languages other than N-S, T&B gives the following: Katla Group: no *p* or *f* (262), Koalib Group: *p* but no *f* (271), Tegali-Tagoi: *f* but no *p* (289). Thus the picture for this compact area—as it was about mid-century before Arabization accelerated alarmingly—was of a mixture of just about all possibilities.

40.4.2. Ehret's reconstructions

Ehret (1989) is an overview of the author's evolving sub-classification of N-S based on phonological and lexical isoglosses. I had thought that a comparison of my reconstructed lexicon with his would be an illuminating exercise, but now I find on inspection that the differences are too great.

In Ehret's scheme, Koman and Gumuz together form a group coordinate with everything else. In my scheme, on the contrary, the two are members of the "Core," rather closely related to East Sudanic and Kado. Ehret's proto-segments are similar to the richest set in Koman, namely that of Twampa (Uduk), which has fivefold manner distinctions in obstruents: voiceless (aspirate, non-aspirate, ejective) and voiced (plain and implosive) as well as six places of articulation: labial, dental, alveolar, palatal, velar, post-velar. My approach is the opposite: I see the luxurious growth in Koman, culminating in Twampa, as an innovation. The situation seems to me to be parallel to that of Afrasian, in which Semitic, especially Arabic, was traditionally taken as a model for the proto-language.

The differences in internal structure are difficult to compare because Ehret's scheme is so detailed, with no less than eight levels against my four or five. But we agree at points, e.g. in grouping E1, E3, and E7 (I adjoin also E5).

40.4.3. Blench's Classification

Roger Blench (1992) has proposed in a recent paper a major restructuring of African language phyla by creating a new Niger-Saharan super-phylum.⁶ This novel proposal makes all of the vast Niger-Congo phylum coordinate to Central Sudanic at the 4th level down in the scheme. It does other startling things, also, e.g. divides up my Core group among Komuz (= Koman + Gumuz), East Sudanic, and Kado in three different parts of the tree.

6. It is only fair to note that Blench refers to his scheme as speculative.

Blench's justification for setting off a group consisting of Central Sudanic vs. Niger-Congo is twofold: vowel-harmony systems involving ATR and presence of labiovelars in the two branches. I find this implausible in view of the existence of a much simpler explanation: diffusion from Niger-Congo into Central Sudanic languages of the area of contact, an area of notable linguistic complexity and multilingualism.

40.5. Main languages of analysis

This is not an exhaustive list of all purported N-S languages. It includes representatives of all known families and sub-families according to my analysis so far as represented in §§ 40.5–40.6. Details of sub-classification are not spelled out herein. For a note on primary language references, see **References**.

40.5.1. *Overall Nilo-Saharan*

- A. Songay
 - Northern Songay
 - Southern Songay: Songay, Zerma, Dendi⁷
- B. Saharan
 - Kanuri-Kanembu, Teda, Tubu, Zagawa-Berti
- C. Maba(n)⁸
 - Maba, Mesalit, Aiki, “Mimi”
- D. For(an)⁹
 - For, Amdang
- E. East Sudanic¹⁰
 - (see §§ 40.5.2, 40.5.3)
- F. Central Sudanic¹⁰
 - (see § 40.5.4)
- G. Berta
- H. Kunama
 - Kunama-Ilit
- I. Koman
 - Twampa, Komo, Opo, Kwama, Gule
- J. Gumuz

7. The situation is much more complicated than shown here (see Nicolai references). The three varieties chosen are representative of Southern Songay; Northern Songay is poorly documented and much influenced by Tuareg.

8. Not to be confused with the Nilotic language Mabaan or Mebaan.

9. “For” and “Fur” are alternative spellings. Amdang is the “Mimi” variety reported by Jungraithmayr.

10. For these large families I refer to my own reconstructions.

- K. Kuliak
 - Nyangi, Ik, Soo
- L. Kado¹¹
 - Kadalla,¹¹ Krongo, Tulishi, etc.

40.5.2. *East Sudanic*

40.5.2.1. *Ek. k-group*¹²

- E1. Nubian
 - Nobiin
 - Dongola (Kenzi)
 - Birgid (extinct?)
 - Midob
 - Dilinj, Kadaru and other "Hill" varieties
- E3. Nera ("Barya"¹³)
- E5. Nyima
 - Nyimang
 - Dinik ("Afitti")
- E7. Tama
 - Tama
 - Erenga-Sungor
 - Merarit
- E9. (see § 40.5.3)

40.5.2.2. *En. n-group*¹²

- E2. Surmic¹⁴
 - Majang (Masongo)
 - Murle, Didinga, Longarim
 - Mursi
 - Me'en (Tishena-Bodi)
 - Kwegu, Muguji, Bale, and other doubtful varieties
- E4. Jebel¹⁵

11. "Kado" is a name for "Kadugli-Krongo" suggested by the late Roland Stevenson. "Kadalla" is the self-name of the Kadugli people.

12. East Sudanic is divided into two sub-groups based on having a (retained) *-k* or an innovated *-n* in the first-person pronoun.

13. The former name "Barya" means 'slave'.

14. The former "Didinga-Murle"; name suggested by Peter Unseth from a common ethnonym in the area.

15. "Jebel" is a local Arabic word for the characteristic rocky hills of Sudan and adjacent territories. This is my name for a group which includes the former "Ingessana" (= "Gaam," meaning the particular hills they inhabited) and three varieties reported by Evans-Pritchard as "Berta languages" (his names given with each in parentheses).

Gaam (“Ingessana, Tabi”)

Aka (“Sillok”)

Kelo (“Tornasi”)

Molo (“Malkan”)

E6. Temein¹⁶

Ronge (“Temein”)

Doni (“Jirru”)

Dese (“Tesei”)

E8. Daju¹⁷

Shatt

Liguri

Nyala-Lagowa

Nyolge (Njalgulgule)

Mongo-Sila

Beygo (extinct?)

40.5.3. *Nilotic*

E9a: West Nilotic

1. North:

Burun, Mebaan, Jumjum

2. Colo (Shilluk)

Anywa (Anuak)

Jur

Turi¹⁸

Bor

Acoli

Kumam (Lango)

Adola

Luo

3. Jieng (Dinka)

Naadh (Nuer)

Atuot¹⁹

E9b: East Nilotic

1. Bari

2. Lotuxo

Maa (Masai)-Ongamo

Toposa

Turkana

16. I use the self-names of these three main “Temein” varieties.

17. “Daju” is a name for the people in question, not accepted by all groups.

18. No Turi data available to me.

19. No significant Atuot data available to me.

Karamojong

Teso

E9c: South Nilotic

1. Omotik (Laamot, extinct)²⁰

2. Datooga (Tatoga)

3. Kalenjin:

Nandi-Kipsikis

Sapiny

Sogoo

Pakot (Suk)

40.5.4. *Central Sudanic*

Fp: Peripheral Group²¹

F1. Miza, Logbara, Lokai, Logo, Avukayo, etc.

F2. Balese, Mamvu, etc.

F3. Mangbetu

F7. Kresh, etc.

F8. Baledha (“Lendu”)

F9. Aja, etc.

Fc: Central Group²¹

F4a. Bagirmi

Sara-Mbai

Sara-Ngambay

Sara-Kaba²²

F4b. Yulu-Binga

F4c. Fongoro

F4d. Shemya (Sinyar)

F4e. Fer(Kara)

F5. Bongo

F6. Modo

Baka

40.6. *Common Nilo-Saharan segments*

See Table 40-3.

20. Not to be confused with “Omotik,” an Afrasian (Afro-Asiatic) language family.

21. The Central Group (Fc) is a genetic group, but Fp may not be—it is merely the rest of Central Sudanic outside of the Fc group. I do not give names to groups, since the ones so far proposed are mainly of the redundant form “Bongo-Bagirmi,” etc. For Fp, see Tucker 1940. For an overview of Fc, see T&B. My classification is partly based on personal communication (1991–92) with Pascal Boyeldieu (who is not responsible for any errors I may have introduced).

22. There are many other “Sara” languages.

Table 40-3. Common Nilo-Saharan Segments

LANGUAGE	LABIALS					DENTALS AND ALVEOLARS						PALATALS				
	Universal: b, m, w					Universal: t, d, l, r, n, s						Universal: y, ɲ				
	p	p'	ɓ	f	v mb	t'	ɖ	ʈ	ɗ	θ	z nd	c	j	c'	š	ʒ ɲj
A. Songay																
(Gao)	(p)			f							z	c=dž	j		š	ž
Dendi	(p)			f							z	c	j			
Zerma	(p)			f							z	c	j			
"Common"				f												
B. Saharan																
Kanuri	(p)			f	mb						z	(c)	(j)		(š)	
Teda	p			f							z	c	j		š	
Tubu	p			f						θ		c	j		š	
Zagawa	p			f?							z	c	j		š	
C. Maba																
Bora-Mabang	p			(v)mb			tr	dr			z nd	c	j		š	ɲj
Mesalit	(p)			f v mb							nd	c	j		š	ɲj
Aiki	p~f		(ɓ)		mb	(d)	(θ)				(z) nd	c	j		š	ɲj
D. For	p~f			f~p							(z)		j			
E. *East Sudanic	p			f-		*t ₂ = t/d	*t ₃ = d/t	*l ₂ = r/l	*l ₃ = l/r	*l ₄ = l-r		(c)	j		š in Ek	
F. *Central Sudanic			ɓ	f v mb			d'	tr?	dr?		(z) nd	c	j		š in Fp	ž in Fp
G. Berta	p ~ p'			f	mb					θ	z nd		j		š	
H. Kunama	(p)			f (v)mb							(z) nd	c=t ^y	j=d ^y		š=s ^y	ɲj
I. Koman																
Twampa	p, p ^h , (p')					t, ɖ, t, t, t ^h , t ^h , Also s'		d ~ d'				c, c ^h	j	c'	š	Also j'
Komo	p'	ɓ									z-					š
Kwama	p'		f			Also s'					(z)					š
Opo		ɓ-				t' d'						c	j	c'	š	
Gule			f			-t' -d'						c=t ^y	j	j'	(š)	
Proto-Koman	p, p ^h ?, p'	ɓ			mb	t, t ^h d, s'									š	
J. Gumuz																
Sai			ɓ	f	mb	t'	ɖ				z nd	c	j	c'	š	ž ɲj
Sese	p	(p' ɓ)	f			t'	d=ɖ				z	c	j	c'	š	
K. Kuliak																
Ik	p		ɓ	f			d				z	c	j			
Soo	p		ɓ				t'	d				c		j'		
Nyangi	p		ɓ				d			l-l'		c		j'		
Proto-Kuliak	p		ɓ			Also dz,				hl	z	c,	j-j',	c'		
L. Kado																
overall			ɓ		mb	Also t, ɖ, t',	d',	t			nd	c	j		š	ɲj
Krongo	p, b = ɓ		f		mb	d=d	t				nd, nt	c, nc			š	

Table 40-3 (Continued)

VELARS AND POST-VELARS						VOWELS							
LANGUAGE	Universal: k, g, ŋ					Universal: i, e, a, o, u							
	k'	g'	ŋg	h	ʔ	ɪ	ɛ	ə	ɔ	ʊ	ai	au	oi
A. Songay													
(Gao)	kp	gb	ŋm	h		Also ii, ee, aa, oo, uu							
Dendi			ŋm	h		Also ii, ee, ɛɛ, aa, ɔɔ, oo, uu							
Zerma				h		Also ii, ee, aa, aiʔ, auʔ, wa, waa, oo, uu							
“Common”				h		Also ii, aa, uu							
B. Saharan													
Kanuri			ŋg	h	(ʔ)	(e), ʌ							
Teda				(h)			ə	Also ii, ü					
Tubu				h	ʔ		ɛ, ɛ	ə	ɔ, ɔ				
Zagawa				h	ʔ	ɪʔ	ɛ		ɔ	ʊʔ			
C. Maba													
Bora-Mabang	Also x		ŋg	h		ɪ	ɛ	ə	ɔ	Also ɤ (ʌ)			
Mesalit			ŋg	(h-x)			ɛ		ɔ	Also (ɤ, ʌʔ)			
Aiki			ŋg	h		ii, ee, ɛ, ɛɛ, aa, ɔɔ, oo, uu, ai, ei, au							
D. For				h-ʔ		Also ii, aa, oo, uu, ai, ei, au							
E. *East Sudanic	Also	g ₂ =g/k				ai oiʔ							
F: *Central Sudanic			ŋ-ŋg		(ʔ)	Also *i-e, *e/i, *a-o, *o-u, *u/o (ai)							
G. Berta	kʰ-k		ŋg	h	ʔ	Also ii, ee, aa, oo, uu							
H. Kunama			ŋg	h		Also ii, ee, aa, oo, uu				ai	au	oi	
I. Koman													
Twampa	k'	(gʔ)		(h)	ʔ					ai			
Komo	(k')			h	ʔ	Also ii, ee, aa, oo, uu				ai	au		
Kwama	k'			h	ʔ	Also ii, ee, aa, oo, uu				ai			
Opo	k'			h	ʔ	Also ei				ai	au		
Gule										ai	au		
Proto-Koman	k' k ^h	g'		h	ʔ								
J. Gumuz													
Sai	kʰ-k		ŋg	h	ʔ	Also ii, ee, ɪ, ɛʔ, aa, oo, uu, ai							
Sese	k'	(g')			(ʔ)	Also ə, ʌ							
K. Kuliak													
Ik	(k')	g'		h		ɪ	ɛ	ə=ʌ	ɔ	ʊ			
Soo		g'			ʔ	ɪ	ɛ	ʌ	ɔ	ʊ			
Nyangi		g'					ɛ		ɔ	ʊ			
Proto-Kuliak	k'	gʼ-g		h		ɪ	ɛ, ɛ		ɔ, ɔ	ʊ			
	Also k ^h , k', k ^w , g ^w , k ^{hw} , k ^{hw}												
L. Kado													
overall			ŋg		(ʔ)	(ɪ)	ɛ	ə	ɔ	(ʊ)ʔ			
	Also k ^w -, g ^w -												
Krongo			ŋk			ii, ɪ, ɪɪ, eeʔ, ʌ, aa, oo, ʊʊ, uu, ʊ							

40.6.1. Notes to Table 40-3.

Phonemes in () are rare and/or marginal. The notation x/y in *East Sudanic and *Central Sudanic (respectively) means x in Ek, y in En; x in Fp, y in Fc.

Notes referring to specific phonemes (columns of the table) are incorporated into the discussion of § 40.3. Those referring to language groups and individual languages (rows of the table) follow.

A: Songay group: even excluding Northern Songay (heavily Berberized), the Songay languages have a great range of diversity in segmental systems; Nicolai (1981: 47, 50) gives as the “fixed system” (called in Table 40-3 “common”): $b, m, w, f, t, d, n, l, s, k, g, y, h, i, ii, a, aa, u, uu, e, o$.

B: Saharan: Kanuri: $p < b, f$; all of p, t , and s occur only before $-i$ and are not phonemic; vowels: at least in some dialects, there are two central vowels: rare mid to high ə , and mid to low a .

B: Teda: $/c/ = [k^{\text{v}}, t^{\text{v}}, t\text{ʃ}]$

B: Tubu: Also: $k^{\text{v}}, x, \gamma, (k^{\text{w}}, g^{\text{w}})$

B: Zagawa: according to Roland Stevenson, both $/t, d/$ (dental) and $/t, d/$ (alveolar) are distinctive

C: Maba Group: Prenasals (mb, nd , etc.) might be treated as sequences; $[k^{\text{w}}]$ and $[g^{\text{w}}]$ occur

C: Bora Mabang: $v < w$?

C: Mesalit: $p < f?$, $v < b?$; z only in Arabic loans

C: Aiki: $\theta \sim s, c \sim \text{ʃ}$

D: No systematic phonological analysis has been carried out on Amdang; $/dz/$ found in Amdang; $y \rightarrow z / _\#$

E: *East Sudanic: $/p/$ found in En, not in Ek (f for p ; see under $/p/$ above); initial only in *East Sudanic: f, w ; final $-d$ and $-y$ and initial r - and n - are rare; final $-n$ alternates with $-\eta$; tendency: $s \rightarrow \text{ʃ} / _\text{i}, e$

F: *Central Sudanic: $/v/ = [v/b]$ (i.e. v in Fp, b in Fc; there are also problematical reconstructed phonemes $I_2 = r/l, I_4 = l \sim r$; also $/*kp \sim gb/?$, $/*ngb/?$; Central Sudanic languages have vowel systems ranging over 5, 7, and 10 basic vowels.

G: Berta: $f \rightarrow \phi, \beta, w / V _\text{V}$; dialectally, $/t/ \sim \theta, p \sim \eta$; $g \sim j / _\text{i}, e, u \sim \alpha$; z and p are found in the most Arabic-influenced dialect (Mayu); intervocalic voicing of s, s' to z and θ to ð is common.

H: Kunama: Dialectally, $h \sim s$

I: Twampa: final $-C \sim C^h$; initial only and rare: s, n, w, p, η, h - rare; also k^h, k^{w}, gw -, etc.

I: Komo: $p^{\text{w}}, t^{\text{w}}$, etc. occur; no initial n, r ; no final $-d, -s$; z - mainly initial; no p

I: Kwama: no p ; no initial $l, r, \eta, \text{ʔ}$; no final $-d, -w, -g, -h$; C^{w} - occurs; p rare

I: Opo: no initial $r, \text{ʔ}$; n - rare; no p ; no final $-b, -f, -d, -t, -s, -g, -k, -h$; C^{w} - occur, all but k^{w} - rare.

I: Gule (nearly extinct and much influenced by Arabic): no final $-j$; $-t'$ - and $-d'$ - medial only; no C^{w} ; $[\text{ʔ}]$ sporadic

J: Gumuz: Sai: $/l, d/, [rr]$ all $\sim r$; also $C^{\text{w}}, k^{\text{ʔ}}, g^{\text{ʔ}}$; $/h/ \rightarrow [x]$

J: Gumuz: Sese: $/j, c, c'/$ are palatal; there are also alveolar and alveo-palatal voiceless series: $/c_1, c'_1/, /c_2, c'_2/$ respectively; initial only: p -, $y \sim \text{zero}, \text{ʔ} / _\text{i}$; often $g \sim g$

K: Kuliak: Ik: Dentals *t*, *d*, and *ŋ*; also in the speech of older persons: *ts'* ~ *tl'*, *l'*, *z* ~ *l'y*

K: Kuliak: Soo: Dentals *s* and *ŋ*

K: Proto-Kuliak: *d* ~ *d'*; no *-dz-*, *z-*, *-l-*, *-j-*; *c'* → *s* / *V*__*V*; *h*- initial only; *h* → *ʔ* morpheme-final.

L: Kado overall: [*z*] < /*s*/; *t* ~ *t*; *ʔ* occurs in pronouns; several centralized vowels; *V* → zero or [-vce] /__#

L: Krongo: /*e*/ and /*ee*/ doubtful

40.6.2. Sources for Table 40-3

A. Songay (all varieties): Nicolai 1981.

B. Kanuri: Hutchison 1981; Teda: LeCoeur 1955; Tubu: Lukas 1953; Zagawa: T&B < Roland Stevenson ms.

C. Maba: Edgar 1989; Mesalit: Edgar 1989; Aiki: Nougayrol 1989.

D. For: Jakobi 1990.

E. East Sudanic: In Table 40-3, *East Sudanic forms from my ongoing work are used. For names of groups and individual languages in East Sudanic, see § 40.5.2-3.

F. Central Sudanic: In Table 40-3, *Central Sudanic forms are from my ongoing work (see Bender 1992 for an earlier version); for names of groups and individual languages in Central Sudanic, see § 40.5.4.

G. Berta: Triulzi, Atieb, and Bender 1976.

H. Kunama: Thompson 1983; Bender ongoing work.

I. All: Bender 1983a; Twampa: Thelwall 1989; Komo: Burns 1947.

J. Sai: Bender 1979; Sese: Uzar 1989.

K. All: Heine 1976, Ehret 1981, Serzisko 1989.

L. All: T&B; Krongo: Reh 1985.

40.7. Reconstructed segmental phonemes

The / is used to describe segments whose distributions are one form in Outliers and another in Satellite-Core, e.g. *r/l* means *r* in Outliers, *l* in S-C; similarly for *e/i*, etc. The ~ is used to indicate possible other segments for which the distributions are unclear, e.g. *L: l-r* represents a possible "fourth" sonorant along with *l*, *r*, and *r/l*, one which might account for modern-day *l*'s and *r*'s in a pattern not yet understood.

The analysis so far indicates that only two syntagmatic positions need to be considered: initial and medial. The root morphemes of N-S are mainly of form CVC with less common CV. However, nearly every CVC appears today with a final -V which is generally unreconstructable, varying from language to language. Also, most roots in contemporary languages appear with prefixes or suffixes so that final -C is rare and initial C is often subject to assimilations.

See further Bender forthcoming.

Table 40-4. Proto-Nilo-Saharan Segmental Phonemes

Consonants			
b	d	j	g
p	t, t ₂	c?	k, kʔ, k ^h ?
f	s	š	h ʔ
m	n	ɲ	ŋ
	l, r, r/l?, L: l-r?		
w		y	
mb	nd	(Satellite-Core only)	
Vowels			
i		u	Other possibilities:
e		o	ii, uu, ee, aa, oo, uu
	a		U: i~u, I: i~e, E: e~a
au	ai	oi	A: a~o, O: o~u
			e/i, o/a, o/u
			ieʔ, eiʔ, iaʔ

References

Since the listing of all primary language sources would take up many pages, they are not given here; these references are limited to general sources and major ones used as sources for § 40.6. See Bender forthcoming and earlier studies for other specific-language references.

Bender, M. Lionel. 1976a. "Nilo-Saharan Overview." In Bender 1976b: 439–83, 512.

———, ed. 1976b. *The Non-Semitic Languages of Ethiopia* (Committee on Ethiopian Studies, Occasional Papers Series, Monograph 5). East Lansing: Michigan State University, African Studies Center.

———. 1979. "Gumuz: A Sketch of Grammar and Lexicon." *Afrika und Übersee* 62: 38–69.

———. 1983a. "Proto-Koman Phonology and Lexicon." *Afrika und Übersee* 66: 259–97.

———, ed. 1983b. *Nilo-Saharan Language Studies* (Committee on Northeast African Studies, Monograph 13). East Lansing: Michigan State University, African Studies Center.

———. 1989a. "Nilo-Saharan Pronouns/Demonstratives." In Bender 1989c: 1–34.

- . 1989b. "Proto-Omotic Phonology and Lexicon." In *Cushitic-Omotic: Papers from the First International Symposium on Cushitic and Omotic Languages*, ed. Marianne Bechhaus-Gerst and Fritz Serzisko, pp. 121–59. Hamburg: Buske.
- , ed. 1989c. *Topics in Nilo-Saharan Linguistics* (Nilo-Saharan Linguistic Analyses and Documentation 3). Hamburg: Buske.
- . 1991. "Sub-Classification of Nilo-Saharan." In *Proceedings of the Fourth Nilo-Saharan Conference*, ed. M. L. Bender (Nilo-Saharan Linguistic Analyses and Documentation 7), pp. 1–35. Hamburg: Buske.
- . 1992. "Central Sudanic Segmental and Lexical Reconstructions." *Afrikanistische Arbeitspapiere* 29: 5–61.
- . 1994. "The Jebel Languages: Lexicon and Phonology." Paper presented at 25th African Linguistics Conference, Rutgers University, New Brunswick, N.J.
- . forthcoming. *The Nilo-Saharan Languages*.
- Bender, M. Lionel, and Alexander Naty. forthcoming. *Kunama-English/English Kunama Dictionary*.
- Blench, Roger. 1992. "Is Niger-Congo Simply a Branch of Nilo-Saharan?" Paper presented at 5th Nilo-Saharan Conference, Nice.
- Burns, S. J. 1947. "Towards a Grammar of the Koma Language." Khartoum: Sudan Interior Mission? (unpublished manuscript).
- Edgar, John. 1989. *Maba Group Lexicon*. Berlin: Reimer.
- Ehret, Christopher. 1981. "Revising Proto-Kuliak." *Afrika und Übersee* 64: 81–100.
- . 1989. "Sub-Classification of Nilo-Saharan: A Proposal." In Bender 1989c: 35–49.
- . ms. "Phonological Reconstruction of Nilo-Saharan (plus extensive list)."
- Ferguson, Charles A. 1976. "The Ethiopian Language Area." In *Language in Ethiopia*, ed. M. L. Bender, J. Donald Bowen, Robert L. Cooper, and C. A. Ferguson, pp. 63–76. London: Oxford University Press.
- Fleming, Harold. 1983. "Kuliak External Relations: Step One." In *Nilotic Studies Part Two*, ed. Rainer Vossen and Marianne Bechhaus-Gerst, pp. 423–78. Berlin: Reimer.
- Goodman, Morris. 1970. "Some Questions on the Classification of African Languages." *International Journal of American Linguistics* 36: 117–22.
- Greenberg, Joseph H. 1963. *The Languages of Africa*. The Hague: Mouton; Bloomington: Indiana University.

- Heine, Bernd. 1976. *The Kuliak Languages (of Eastern Uganda)*. Nairobi: East African Publishing House.
- Hutchison, John. 1981. *The Kanuri Language*. Madison: University of Wisconsin, African Studies Program.
- Jakobi, Angelika. 1990. *A Fur Grammar* (Nilo-Saharan Linguistic Analyses and Documentation 5). Hamburg: Buske.
- Le Cœur, Ch., and M. Le Cœur. 1955. *Grammaire et textes Teda-Daza* (mémoires de l'Institut Français d'Afrique Noire 46). Dakar.
- Lukas, Johannes. 1953. *Die Sprache der Tubu in der zentralen Sahara* (Deutsche Akademie der Wissenschaften zu Berlin, Institut für Orientalforschung Veröffentlichung 14). Berlin: Akademie Verlag.
- Newman, Paul, and Roxanna Ma. 1977. "Chadic Classification and Reconstructions." *Afroasiatic Linguistics* 5/1: 1–42.
- Nicolai, Robert. 1978. *Les dialectes du Songhay* (SELAF 85–86). Paris: Centre National de la Recherche Scientifique.
- . 1990. *Parentés linguistiques (à propos du songhay)*. Paris: Éditions du Centre National de la Recherche Scientifique.
- Nougayrol, Pierre. 1989. *Les Langue des Aiki*. Paris: Geuthner.
- Reh, Mechtild. 1985. *Die Krongo-Sprache* (Kölner Beiträge zur Afrikanistik 12). Berlin: Reimer.
- Serzisko, Fritz. 1989. "The Kuliak Languages: A Structural Comparison." In Bender 1989c: 385–404.
- Stevenson, Roland C. ms. [on Zagawa. Cited in T&B; seen by MLB in 1991 or 1992].
- Thelwall, Robin. 1983. "Twampa Phonology." In Bender 1983b: 323–35.
- Thompson, E. David. 1983. "Kunama: Phonology and Noun Phrase." In Bender 1983b: 280–322.
- . 1989. "Kunama Verb Phrase." In Bender 1989c: 305–46.
- Triulzi, Alessandro, Atieb A. Dafallah, and M. Lionel Bender. 1976. "Berta." In Bender 1976b: 513–32.
- Tucker, Archibald N. 1940. *The Eastern Sudanic Languages*, vol. 1. London: International African Institute.
- Tucker, Archibald N., and Margaret A. Bryan. 1956. *The Non-Bantu Languages of North-Eastern Africa*. London: Oxford University Press.
- . 1966. *Linguistic Analyses: The Non-Bantu Languages of North-eastern Africa*. London: Oxford University Press.
- Uzar, Henning. 1989. "Studies in Gumuz: Sese Phonology and TMA System." In Bender 1989c: 347–83.

5 Niger-Congo Languages

This page intentionally left blank.

Swahili Phonology

Ellen Contini-Morava

University of Virginia

41.1. Social and historical background

Swahili is a Bantu language, more specifically a member of the Sabaki subgroup of North East Coast Bantu. It has been suggested that the ancestor of the modern dialects was spoken in an area along the East African coast somewhere between the Webi Shebelle River in what is now Somalia and the Tana River, in Kenya (Nurse and Spear 1985: 46; Nurse and Hinnebusch 1993 [henceforth N & H]: 490–96). On the basis of archaeological and linguistic evidence, the emergence of Proto-Swahili as distinct from the other Sabaki languages has been tentatively dated to approximately A.D. 700–800 (N & H 493).

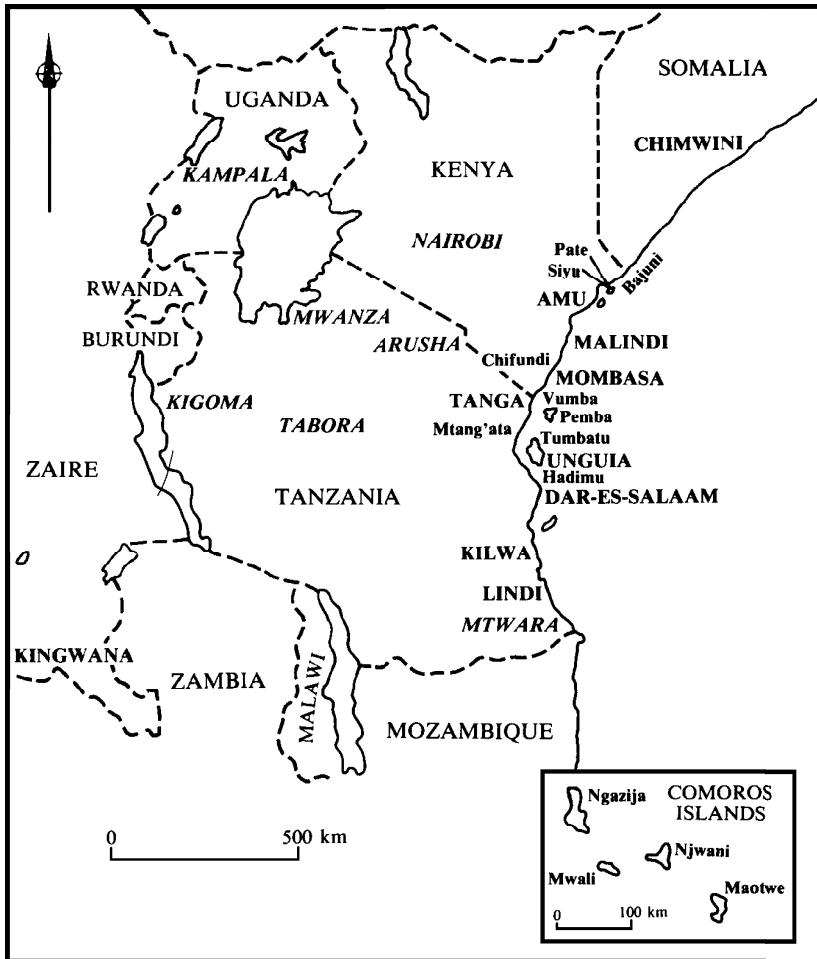
The earliest Swahili-speaking communities had a mixed herding, farming, and fishing economy, but the aspect of their economy that became a defining characteristic of Swahili civilization was the lucrative trade networks that the Swahili developed, linking the African mainland with traders from southern Arabia and the Indian Ocean. Major exports included ivory, ambergris, gold, and slaves. Aided by the seasonal monsoon winds, the maritime Swahili established a string of settlements along the East African coast and adjacent islands, eventually stretching all the way from southern Somalia to northern Mozambique, northern Madagascar, and the Comoro Islands. Between A.D. 1100 and 1500 the Swahili dominated trade between the African interior and the Indian Ocean, a hegemony that was interrupted in the 16th century by invasions from the south by the Portuguese and from the north by Orma and Somali raids. During the 18th and 19th centuries, under the rule of Omani Arab sultans who ousted the Portuguese and established a court in Zanzibar, the Swahili regained their position as mediators of trade between the mainland and the Indian Ocean, establishing caravan routes and trading outposts between Zanzibar and places as far west as Uganda, Burundi, and Zaire. In this way the language was carried into the African interior, where a number of other languages were spoken, mostly also Bantu,

and Swahili became a lingua franca outside its native region, especially in the urban centers that developed from the trading posts.

The usefulness of Swahili was recognized by the European colonial administrators who eventually controlled East Africa, and during the period of European colonialism, from the late 19th century to the mid 20th century, Swahili was used as an administrative language and a medium of education. During the 1930s a standardized variety was developed based on the dialect of Zanzibar City (Whiteley 1969). Since independence, Standard Swahili has been promoted vigorously in Tanzania, where it is the national language, and to a lesser extent in Kenya, where it also enjoys official status alongside English. By now the native speakers of traditional Swahili ancestry, perhaps 2 million, are far outnumbered by the people in Kenya, Tanzania, Uganda, and eastern Zaire who learn it as a second or third language and use it as a lingua franca, especially in urban centers where it is becoming the mother tongue of the emergent working class. These latter speakers are estimated to be as many as 50 million (Wald 1987: 995).

Because of its role as a trade language, Swahili has come into contact with a number of other languages, of which the most influential has been Arabic. Islam has been an important component of Swahili identity for several centuries, although it is uncertain whether widespread conversion to Islam co-occurred with or post-dated the development of the major Indian Ocean trade networks (Nurse and Spear 1985: 95 ff.). Swahili was first written in Arabic script, in which there are manuscripts dating back to the early 18th century. (A Roman alphabet was introduced during the colonial period and is now standard, although some Swahili-speaking Muslims continue to use Arabic script, especially in private correspondence.) The cultural importance of Islam is reflected in the large number of loanwords from Arabic; indeed this fact misled some early scholars to describe Swahili as a "mixed language," a view that persists today among many East Africans. In fact, the grammatical structure and the core vocabulary of Swahili are unambiguously Bantu, and the majority of Arabic loanwords entered the language relatively recently; most date back only as far as the period of Omani Arab domination in the 18th and 19th centuries (N & H 315). Although the grammatical structure of the language has been unaffected by its contact with Arabic, the phonological system has absorbed some Arabic sounds along with the borrowed vocabulary, a point to be discussed below.

There are a number of dialects of Swahili (see map). As in most cases of dialect classification, conflicting criteria for subgrouping have led researchers to somewhat different results. The most recent and most comprehensive



Map 4. The Swahili-speaking Area (reprinted from Wald 1987: 993, with permission). Urban dialects are capitalized. Urban dialects in italics are urbanizing areas where non-traditional dialects of Swahili are currently developing. Nations are in large capitals.

treatment of Swahili historical phonology and morphology to date is N & H (for dialect subgrouping see especially chapter 6, section 6ff.). On the basis of shared innovations, especially phonological, N & H recognize a Northern Dialect group that includes Chimwini, Bajuni, Siyu, Pate, Amu, Malindi, Mombasa, and possibly Chifundi (see map; N & H 504). The rest of the dialects they tentatively assign to a Southern group, though they admit

that this group is less cohesive and indeed that lexicostatistic findings conflict with other criteria for subgrouping (N & H 512).

As mentioned earlier, Standard Swahili is based on the dialect of Zanzibar City, part of the Southern group (see Batibo 1989 for discussion of differences between Zanzibar and Standard Swahili, especially that of Dar es Salaam, the capital of Tanzania). Through the influence of the media and education this dialect is significantly affecting the smaller, local dialects, especially in Tanzania (Nurse and Spear 1985: 61; N & H 32). In this chapter I will focus on Standard Swahili, giving information on the major phonological differences between it and the other dialects in the historical section. In the interest of space I will also confine myself to word-level phonology. For a detailed treatment of stress and intonation in phrases and clauses, see Maw and Kelly 1975.

41.2. Consonants

Table 41-1. Standard Swahili Consonants

	Bilabial	Labiodental	Dental	Alveolar	Palatal	Velar	Laryngeal
Stop	p			t		k	
	p ^h			t ^h		k ^h	
	b			d	ʃ	g	
	m ^b			ⁿ d		^ŋ g	
Fricative		f	(θ)	s	š	(x)	h
		v	(ð)	z		(ɣ)	
Affricate					č		
					č ^h		
					ǰ		
Nasal	m			n	ɲ	ŋ	
Lateral				l			
Trill				r			
Semivowel	w				y		

The consonant segments of Standard Swahili are shown in Table 41-1.

/p/ [p] /t/ [t] /k/ [k] Inherited from Common Bantu. These phonemes may be aspirated in stressed syllables (see § 41.2.1).

/p^h/ [p^h] /t^h/ [t^h] /k^h/ [k^h] Note: the aspirated consonant series /p^h t^h k^h ɕ^h/ have a restricted distribution: they are the historical result of deletion of a preceding *n, usually the prefix of Bantu noun class 9/10 (see § 41.2.1).

/b/ [b] voiced imploded bilabial stop. Reflex of Proto-Sabaki *W/_*j (N & H 134).

/d/ [d] voiced imploded apico-alveolar stop. A dental stop [ɖ] in northern dialects (see § 41.7.1).

/ɟ/ [ɟ] voiced imploded medio-palatal stop. Reflex of *j (exact feature specification uncertain, N & H 110); has changed to /y/ in some northern dialects (Nurse 1982: 78).

/ɡ/ [ɡ] voiced imploded dorsovelar stop.

/ᵐb/ [ᵐb] prenasalized voiced bilabial stop.

/ᵐd/ [ᵐd] prenasalized voiced apico-alveolar stop. Rhotacized /ᵐdr/ in northern dialects, perhaps to keep this phoneme distinct from the prenasalized voiced dental that developed from Proto-Bantu *ndj/ngi/nj (see § 41.7.1).

/ᵐɡ/ [ᵐɡ] prenasalized voiced dorsovelar stop.

/ɕ/ [ɕ] Reflex of Proto-Sabaki *ky (N & H 74) and *c (N & H 83). The reflex of *c is dental /ɕ/ in northern dialects; in Chimwini /t/ and /t̪/ are separate phonemes.

/ɕ^h/ [ɕ^h] Like the aspirated stops, restricted in its distribution (see § 41.2.1).

/ɲ/ [ɲ] prenasalized voiced apico-alveolar affricate. In northern dialects, a prenasalized dental stop /ᵐɖ/. Note: the prenasalized stops /ᵐb, ᵐd, ᵐɡ/ are exploded, unlike the voiced stops; /ᵐj/, also exploded, differs in point of articulation from the voiced stop /j/ (see § 41.2.2).

/f/ [f] Reflex of Common Bantu *p, *t, or *k/_*ɥ, or *p/_*j (N & H 114, 117).

/θ/ [θ] voiceless interdental fricative. Occurs only in Arabic loanwords (see § 41.2.4).

/s/ [s] Reflex of Common Bantu *t/_*j (N & H 114).

/ʃ/ [ʃ] May be palatalized to [ç] before /i/. Reflex of Common Bantu *k/_*j (N & H 115).

/x/ [x] Only in Arabic loanwords (§ 41.2.4).

/h/ [h] Rare in inherited lexicon, where it is usually a reflex of *nk. Also occurs in loanwords.

/v/ [v] Reflex of Common Bantu *b, *l, or *ɡ/_*ɥ, or *b/_*j (N & H 114, 117).

/ð/ [ð] Only in Arabic loanwords (§ 41.2.4).

/z/ [z] Reflex of Common Bantu **d*, **g* / _ **j*. Replaced by [ð] in some northern dialects. Note: In Chifundi /f v s z/ are affricated to [pf bv ts dz] (Lambert 1958: 16, quoted in N & H 571).

/ɣ/ [ɣ] Only in Arabic loanwords (§ 41.2.4). Note: in the speech of non-Muslims and second-language speakers, /θ ð x ɣ/ may be replaced by /s z h ɡ/ respectively.

/m/ [m] /n/ [n] /ɲ/ [ɲ] /ŋ/ [ŋ] Note: /m n ŋ/ may be syllabic and may carry word stress (on penultimate syllable), cf. § 41.4.2 and § 41.5.

/l/ [l]

/r/ [r] voiced apico-alveolar trill; variably pronounced as a flap (Polomé 1967:44). Note: the *r*/*l* phonemic contrast is relatively recent, and there are many alternations between *r* ~ *l* in morphemes of Bantu origin (cf. § 41.2.3).

/w/ [w] (in northern dialects usually a labiodental approximant [N & H 568]).

/y/ [y] Variably pronounced as [ɜ] in Tikuu (N & H 568).

41.2.1. *Aspiration*

Most grammars of Swahili recognize a voiceless aspirated series /p^h t^h č^h k^h/ alongside the other voiceless consonants (e.g., Myachina 1960 [1981]: 5; Polomé 1967: 38–40), on the grounds that minimal pairs exist that are distinguished only by aspiration, e.g., [paa] ‘roof’ vs. [p^haa] ‘gazelle’. Aspiration is a complicated issue in Swahili, and there are conflicting reports about which speakers consistently make the distinction. Even in the speech of those who do make the distinction, the aspirated consonants have a restricted distribution: they occur only in contexts where a voiceless stop or affricate was historically preceded by a nasal which has been lost (Steere 1870: 11; Tucker and Ashton 1942: 85). Thus [p^haa] derives from **ni-+paa*, with loss of /i/ and assimilation of the nasal of the prefix, and possibly nasal devoicing as intermediate steps (Hinnebusch 1975; N & H 159); compare the cognate Zulu form *impala* ‘gazelle’. Most often this “philological” aspiration (Tucker and Ashton’s 1942 term) appears on the initial consonant of Bantu noun class 9/10 and adjectives agreeing with nouns of this class, where the nasal prefix has been lost, as in the example just given. But aspiration can also occur word-medially, where the nasal used to be part of the stem, e.g., [nuk^ha] ‘smell’ < Proto-Sabaki **nunka*, N & H 601. Complicating this picture is the fact that at least some speakers tend to aspirate /p t č k/ in stressed syllables, regardless of their “philological” status (Tucker and Ash-

ton 1942: 88; Polomé 1967: 41; see also Harris 1951: 99); conversely, even “philological” aspirates may be unaspirated in unstressed syllables (Tucker and Ashton 1942: 89). Furthermore, there is some dispute as to whether the aspirated/unaspirated distinction is gradually being lost in southern Swa. (including Standard), due to its low functional load and possibly the influence of non-native speech; aspiration is also not indicated in the writing system. Polomé (1967: 40) claims that the distinction is no longer consistently made in Zanzibar City or on the adjacent coast, and Wald (1987: 997) apparently concurs, but Engstrand and Lodhi (1985: 79) argue that they found phonemic aspiration to be alive and well in the speech of native speakers from this area; Nurse (pers. comm.) reports distinctive aspiration on the northern Tanzanian coast and parts of Zanzibar, and Hinnebusch (pers. comm.) has observed it in Zanzibar City. N & H include a set of aspirated consonants in their chart of phonemes of Zanzibar dialect (p. 571). The northern dialects regularly make this distinction, and some even have aspirated prenasalized voiceless stops (N & H 568). I have included the aspirated set in the consonant inventory because their occurrence does not seem to be predictable by phonological rules,¹ but it should be noted that they have a somewhat marginal status within the Swahili consonant system.

41.2.2. *Voiced and prenasalized stops*

The non-prenasalized voiced stops /b, d, ɟ, g/ are implosive in all Swa. dialects (N & H 568). Some reference grammars list [b, d, ɟ, g] as (ex)plosive allophones of these after homorganic nasals (e.g., Polomé 1967: 41), thus treating [mb], [nd], [nɟ], [ŋg] as consonant clusters rather than unit phonemes. Some remain agnostic on this point (Tucker & Ashton 1942: 91 call [mb] etc. “nasal compounds,” without being specific as to their phonemic status, but Ashton 1944: 3–4 includes “nasal compounds” *mb, nd, nj, ng* as well as *mv* and *nz* in her chart of Swahili consonants). More recent studies argue that an independent prenasalized stop series should be reconstructed for Common Bantu, on the grounds that these sounds show a separate historical development from the non-prenasalized consonants in the various Bantu languages (Wald 1987: 995; N & H 1993: 145). Synchronic evidence in favor of a separate prenasalized series for Swa. includes the fact that not

1. Polomé argues that aspiration can be predicted by a morphophonemic rule that replaces /n/ with aspiration of the following voiceless stop or affricate in the context #_V (1967: 182n.), but this would not account for stem-medial aspiration. Nor could one write a phonological rule replacing /n/ with aspiration in all contexts, because sequences of /n/ + C occur in loanwords such as /senti/ ‘cent’ (< English).

all instances of homorganic nasal + voiced stop are pronounced with an explosive stop: Tucker and Ashton (1942: 100) report [mbovu] ‘rotten’, adjective stem /bovu/ ‘rotten’ following *m*-prefix of Noun Class 3; [ᵐbovu], same adjective stem following noun prefix of Class 9/10 (confirmed by personal observation of present author).

Whenever a stop is imploded after a homorganic nasal, there is a morpheme boundary between the nasal and the stop (the initial *m*- is a noun class prefix in both the examples just given). But the presence of a morpheme boundary does not guarantee that the stop will be imploded: stem-initial voiced stops are always exploded after the nasal prefix of noun class 9/10, but never after the *m*- of other noun classes. Also, the prenasalized stops pattern quantitatively more like unit consonants than like consonant clusters: they are more frequent within morphemes (nonborrowed lexicon) than all other clusters put together; in intervocalic position their frequency rivals or surpasses that of voiceless stops, voiced stops, and simple nasals (Contini 1969). Even though they are restricted in their distribution within morphemes, being mainly confined to intervocalic position, it seems reasonable to regard /ᵐb, ᵐd, ᵐj, ᵐg/ as unitary prenasalized consonants rather than consonant clusters. I know of no compelling arguments for treating [mv] or [nz] as unitary phonemes, as Ashton’s (1944: 3) analysis implies.

41.2.3. /r/ and /l/

Most Bantu languages do not distinguish between /r/ and /l/; no such distinction is reconstructed for Common Bantu (Guthrie 1967–71, 1:61–62). With a few exceptions, most of the Sabaki languages inherited /l/ as a reflex of Common Bantu **l* (N & H 312). A few developed [r] as an allophone of /l/ before or after front vowels; in some cases /l/ was replaced by /r/ (N & H 103). The phonemicization of the *r/l* distinction in Standard Swa. was promoted by the large number of loanwords containing /r/ (Polomé 1967: 44–45; N & H 312). Nevertheless there is still a certain amount of variation between these phonemes: the *Standard Swahili–English Dictionary* (Johnson 1939) lists alternative forms with /r/ and /l/ for several lexical items, e.g., *loga* ~ *roga* ‘bewitch’ (see also Tucker and Ashton 1942: 95).

Although Swa. /l/ was inherited from Proto-Sabaki, the original consonant was not preserved in all contexts. **l* was lost before non-front vowels, especially in unstressed position, in most Swa. dialects, yielding sequences of vowels (N & H 100): **inula* ‘rise up’ > *inua*; **mulo* ‘digging stick’ > *muo*; **kulu* ‘big, mature’ > *kuu*. This process is no longer productive; sequences of

/l/ + non-front vowel occur in modern Standard Swa., especially in loanwords: *-lala* ‘sleep’; *hilo* ‘that’ (demonstrative of Class 5); *lulu* ‘pearl’ (from Arabic).

41.2.4. Borrowed consonants

Swa. /θ/, /ð/, /x/, and /ɣ/ were borrowed from Arabic, and occur only in loanwords. There is some variation among Swa. speakers in the pronunciation of these loanwords. The borrowed phonemes are most likely to occur in the speech of Muslim native speakers from the coast, who have had some exposure to Arabic, and for whom pronunciation of these sounds as closely as possible to the Arabic model is a matter of prestige. In the speech of non-Muslims and non-native speakers, these phonemes are generally replaced with /s/, /z/, /h/, and /g/ respectively (N & H 312).²

In highly formal speech, such as a recitation in a mosque, an even more Arabized pronunciation of Arabic loanwords may be encountered, including pharyngealized-velarized (“emphatic”) pronunciations of /t/, /s/, and /ð/ with the appropriate allophonics of the following vowel, velarized [ɬ], [q] for /k/, dental [ɟ] for /t/, use of pharyngeal fricatives [ħ] and [ʕ] (Polomé 1967: 45–46; Tucker 1946: 861–67), and geminated consonants (N & H 567; Tucker and Ashton 1942: 99). These do not normally occur in casual speech.

41.3. Vowels

A symmetrical 7-vowel system has been reconstructed for Proto-Bantu (Guthrie 1967–71, 1:46), including “super-close” **i* [i] and **u* [u]. Proto-Bantu also distinguished between long and short vowels (cf. Meeussen 1979) and between high and low tone. N & H reconstruct a seven-vowel system for Proto-Sabaki, although most of the modern Sabaki languages, including Swa., have a 5-vowel system in which the “super-close” vowels have merged with /i/ and /u/ (N & H 61, 175ff.). Where pitch is still distinctive in the Sabaki group, the systems have become pitch-accent systems, defined as “distinctive but restricted use of pitch, with no more than one pitch contrast per morpheme” (Philippson 1993: 249). Tone is not phonemic in Standard Swa., though it still is in some of the outlying dialects (see Philippson, *ibid.*). Contrastive vowel length has been lost in all Sabaki languages except Chimiwini, a northern Swa. dialect (N & H 465); however, see § 41.3.2.

2. N & H (312) also point out that loanwords have reinforced the functional load of /h/, /r/, /b/, /d/, and /s/, which originally had a much more restricted distribution than they do now.

41.3.1. *Vowel quality*

Table 41-2. Standard Swahili Vowels

i					u
		[ɪ]		[ʊ]	
		[e]		[o]	
ɛ			ə		ɔ
		[a]		[ɒ]	
		ɑ			

In Swa. /i/ and /u/ have allophonic [ɪ] and [ʊ] in unstressed syllables; conversely, /e/ and /o/ have slightly higher allophones in unstressed syllables and before nasals and prenasalized stops (Tucker and Ashton 1942: 80–82; Polomé 1967: 46–47).³

/a/ has allophonic [ə] in unstressed position, [ɒ] after /w/, and [a] after /y/ and coronals (Tucker and Ashton 1942: 82–83; Polomé 1967: 47–48).

All vowels may be slightly lengthened in stressed syllables (normally the penultimate syllable, see § 41.5); before nasals and prenasalized stops (Polomé 1967: 47); and may be nasalized in nasal-initial syllables or before /nz/ (Tucker and Ashton 1942: 84).

In Arabized speech styles, [ay], [ey], and [aw] in Arabic loanwords such as *šawri* ‘intention’ may be pronounced as diphthongs, but there is a tendency either to give syllabic value to each part of the diphthong, resulting in a dissyllabic pronunciation ([ʃauri]), or to coalesce the diphthong into a monophthong, e.g., [ʃaix] ~ [ʃeix] ~ [ʃe:x] ‘chief’ < Ar. *šaix* (Tucker 1946: 870; Polomé 1967: 48).⁴

41.3.2. *Vowel length*

No reference grammars set up a contrast between long and short vowels for Standard Swa., reporting instead subphonemic vowel lengthening in

3. Speakers vary on this point (Tucker and Ashton 1942: 47; Polomé 1967: 47). In fact, Petryankina (1964: 120) reports a lowered rather than raised variant of /b/ in unstressed and pre-nasal contexts. Her informants came from a number of different geographic areas, and included some second language speakers, which may account for the discrepancy.

4. Alan Kaye (pers. comm.) points out that Omani Arabic, the source for most Arabic loanwords in Swa., was a colloquial dialect in which many of the diphthongs of Classical Arabic had already been monophthongized. This suggests that the pronunciation of these sounds as diphthongs in formal speech is due to influence from the classical language—the holy language of the Qurʾān—rather than stemming from familiarity with spoken Arabic.

stressed (penultimate) syllables, as mentioned in § 41.3.1, and treating other instances of lengthened vowels as doubled vowels, each counted as a separate syllable (Tucker and Ashton 1942: 163ff.; Polomé 1967: 57–58). These latter are written as doubled vowels in Standard Swa. orthography, e.g., *kaanga* [ka'aŋga] 'fly' vs. *kanga* ['kʰaŋga] 'guinea fowl'. However, Batibo (1990) argues that vowel length is contrastive even in unstressed syllables in Standard Swa., and that there is a distinction between long vowels and geminated vowels. The following examples are cited (Batibo 1990: 60, 67; Batibo's notation):

<i>kulaani</i>	[kulaá.ni]	'to curse'
<i>kulana</i>	[kulâ:na]	'to eat each other'
<i>kulima</i>	[kulí.ma]	'to cultivate'
<i>kanga</i>	[kâ:nga]	'guinea-fowl'
<i>kaanga</i>	[kaâ:nga]	'fry'
<i>majini</i>	[maɟi:ni]	'genies'
<i>majini</i>	[ma.ɟi:ni]	'in the water'

In an experiment with a mixed group of native and second-language speakers of Swa. at the University of Dar es Salaam, Batibo found that the subjects accurately distinguished between pairs of homographs on the basis of his pronunciation of the words in isolation with greater than chance, though not 100%, frequency. A follow-up study with four native speakers from Zanzibar confirmed these results. The results are certainly puzzling: if length is indeed contrastive, why did the percentage of correct guesses average only 75–85%? But if length is not contrastive, why would the results not be random? Batibo argues that the length distinction is in fact disappearing due to low functional load and low perception (1990: 61). Because this distinction seems to be marginal at best, I have not included it in the vowel inventory.

41.4. Syllable structure

41.4.1. Syllable types and consonant clusters

The preferred syllable type of Swa. is CV or V; in a count of all monosyllabic and disyllabic roots in the *Standard Swahili–English Dictionary* (Johnson 1939), excluding loanwords, a total of 1805 roots, 1274, or 70.5%, were CVCV or VCV (Contini 1969).⁵

5. The purpose of the count was to look at phonotactics within the morpheme. Prefixes were therefore excluded. However, the verb stem final indicative suffix *-a* was counted as part of the root, in order to make it clear that morpheme-final consonants are never word-final in Swa.

Virtually all consonant clusters have either a nasal as first member or a glide as last member; /y/ occurs only after labial consonants (cf. Kelly 1989: 29), whereas /w/ is not so restricted. The following are found in non-borrowed lexis: *pw, tw, kw, bw, dw, gw, mw, mv, nz, vy, fy, sw, py, my, nw, ^mbw, ^ŋgw, ^ɲw, mvy, lm, št*. The last two occur in only one root each: *falme* ‘king’ and *štua* ‘startle’. Both of these clusters are evidently the result of deletion of an intermediate vowel, /u/ and /i/ respectively.

41.4.2. Syllabic nasals

/m/, /n/, and /ŋ/ and the nasal portions of /^mb ⁿd ^ŋg/ are sometimes syllabic in Swa. In these cases, usually prefixes, an original vowel following the nasal has been lost. Nasals are syllabic under the following circumstances:

(a) if one of the homonymous *m*-prefixes (historically **mu*), including the prefixes of noun classes 1 and 3, the 3rd person singular animate verbal object prefix, and some others, precedes a consonant. Examples (TAM = tense-aspect-modality):

m̥toto ‘child’ (Class 1)

m̥buyu ‘baobab tree’ (Class 3)

ninampig’a ‘I hit him’ (*ni* = 1sing. subject prefix, *na* = TAM prefix, *m* = 3sing. object prefix, *-pig’a* = verb stem ‘hit’)

(b) if the nasal precedes a monosyllabic stem, i.e., a stem consisting of a single vowel preceded by one or more consonants. In such cases the nasal also takes the (penultimate) word stress. Examples:

m̥bwa ‘dog’ (< Class 9 prefix **ni* = *bwa*, with loss of vowel and assimilation of nasal)

ŋ̥ge ‘scorpion’ (same class as *mbwa*)

ŋ̥či ‘country’ (same class as *mbwa*).⁶

(c) Optionally, the /i/ of one of the {ni} morphemes (1st person singular subject prefix; copula) may be deleted, leaving behind a syllabic [ŋ], e.g., *ni-nazo* ~ *ŋnazo* ‘I have them’; *ni nini* ~ *ŋnini* ‘what is it’ (Tucker and Ashton 1942: 100).

(d) Optionally, the sequence /mu/ > [m] / _C stem-medially; e.g., *amuka* ~ *aṃka* ‘wake up’ (Tucker and Ashton 1942: 103).

6. Note that monosyllabic stems are an exception to the historical rule that deleted the nasal prefix **ni* before voiceless consonants, replacing it with aspiration in the case of stops and affricates (see § 41.2.1). Tucker & Ashton (1942: 99) report that voiceless stops and affricates are aspirated in monosyllables anyway, even though the motive for aspiration (nasal loss via nasal devoicing) is not there.

41.4.3. Borrowed consonant clusters

Consonant clusters in loanwords that do not conform to Swa. patterns may be broken up with epenthetic vowels. Usually /u/ is inserted after labial consonants, /i/ otherwise. For example:

[buraʃi] 'brush' (< English)

[siliŋgi] 'sling' (< English)

Sometimes a vowel is inserted that matches the vowel in the preceding or following syllable:

[bahari] 'sea' (< Arabic *baħr*)

[foroða] 'customs house' (< Arabic *furða(t)*)

(examples from Tucker 1946–47: 857–58).

41.4.4. Vowel harmony

The vowels of most verb derivational suffixes (causative, stative, “prepositional,” etc.) harmonize with the height of the last vowel in the verb stem: if the verb stem has a mid vowel (/e/ or /o/), so does the vowel of the suffix, and if the verb stem has a non-mid vowel (/a/, /i/, or /u/), the suffix vowel is high. Examples:

vuŋja 'break' *vuŋja* 'break for'

taka 'want' *takia* 'want for'

soma 'read' *somea* 'read for'

leta 'bring' *letea* 'bring for'⁷

Kelly (1989: 27) points out that vowel harmony is pervasive not only between verb stem and suffix, but within verb stems as well, though not apparently within noun stems. He finds that in disyllabic verb stems in Mombasa Swa. the vowels are either central (/e/, /o/) or non-central (/i/, /u/, /a/). Although his study refers only to Mombasa Swa., his observation seems to be valid for Standard Swa. as well: of 168 disyllabic verb stems (non-loans) in Johnson (1939), 150, or 89%, show this type of harmony. Furthermore, the generalization holds even more strongly for N & H's reconstructions of Proto-Sabaki: of 174 verb stems containing more than one vowel, 169, or 97%, show harmony.⁸

7. Space does not permit comprehensive discussion of the morphophonemics of the verb derivational suffixes here. For details, see the reference grammars; Tucker and Ashton 1942: 161ff.; Port and Shepardson 1982.

8. I excluded the 22 stems that N & H reconstructed with “super-high” *j and *ɥ, because I was not sure what to predict about their distribution vis-à-vis the “normal” high vowels /i/ and /u/. In case the reader is curious, the super-high vowels may co-occur with either high or low vowels, but apparently never with a mid vowel. I should also point out that some of N

41.5. Stress

The general rule is that primary stress is on the penultimate syllable (which may be a syllabic nasal), in polysyllabic words. The few monosyllables in Swa. express grammatical meanings (e.g., the copula *ni, na* ‘and/with’, etc.) and are not stressed. Some polysyllabic loanwords are exceptional and have stress on the antepenultimate: [amérika] ‘America’, [itália] ‘Italy’. In fact there is even a minimal pair differentiated only by stress: [barábara] ‘exactly’ and [barábára] ‘highway’, which leads Polomé (1967: 51) to regard stress as phonemic. Vitale (1982) observes that not all loanwords are treated alike in Swa. Some, like those just cited, are exceptions to the penultimate stress rule. Some show variable stress placement: [áfrika] ~ [afríka] ‘Africa’, [lázima] ~ [lazíma] ‘necessity’ (< Arabic). Yet others have been assimilated to the normal Swa. pattern: [bulangéti] ‘blanket’. He suggests differentiating between “historical loanwords” and “phonological loanwords”; the latter either are not assimilated (like [áfrika]) or variably assimilate (like [lázima] ~ [lazíma]), and can be marked as such in the lexicon. Given the rarity of non-penultimate stress, an analysis along these lines seems reasonable, but it would have to be modified to accommodate two exceptions to the penult stress rule in morphemes of Bantu origin, pointed out by Maw and Kelly (1975: 3). These are the auxiliary verb/aspectual marker *-isha* ‘finish’, often shortened to [ʃa], used to reinforce completedness of action, e.g. /nilišámwona/ ‘I saw her (already)’; and the negative first person singular prefix *sí*, sometimes stressed “for purposes of prominence” as in /síwezi kukubalia/ ‘I cannot agree with you’ (Maw and Kelly 1975: 3).

41.6. Morphophonemic processes

The phonological phenomena discussed in this section are pervasive but not automatic: they are restricted to certain groups of morphemes and sometimes to individual lexical items. Principal sources for this section are Polomé (1967: 59ff.) and Tucker and Ashton 1942: 163ff.).

& H’s verb reconstructions include what may be fossilized forms of harmonizing derivational suffixes (for example, **-cangamuk-* ‘be pleased’ and **-elek-* ‘face towards’ seem to contain the stative suffix **-ik* ~ *-ek*, which itself predates Proto-Sabaki; cf. N & H 370). This would inflate the figures for harmony to some extent, but non-harmonizing stems are still extremely rare.

41.6.1. Vowel processes

41.6.1.1. Contraction

Prefix-final /a/ is often deleted before /a/, /e/, and /o/. Examples (from Polomé 1967: 59-60):

/wana/	‘children’	< {wa} Noun Class 2 prefix + {ana} N stem
/mema/	‘good’	< {ma} Class 6 concordial prefix + {ema} Adj. stem

/pote/	‘everywhere’	< {pa} locative prefix + {ote} Pro. stem
--------	--------------	--

Affix-final /a/ coalesces with initial /i/, yielding /e/. Examples:

/meko/	‘ovens’	< {ma} N. Class 6 prefix + {iko} N stem
/someni/	‘read! (pl.)’	< {som} V stem + {a} indicative suffix + {ini} 2pl. suffix

41.6.1.2. Glide formation

Prefix-final /i/ and /u/ may become non-syllabic before vowel-initial morphemes:

/wataka/	‘you (sg.) want’	< {u} 2sg. subject prefix + {a} TAM marker + {taka} V stem
/yote/	‘all’	< {i} Noun Class 9 concordial prefix + {ote} Pro. stem

41.6.2. Consonant processes

Processes affecting nasals and aspiration have already been discussed. Other consonant processes include the following.

41.6.2.1. /n/ + liquid

The sequence /n/ + liquid becomes prenasalized stop /ⁿd/ (Polomé 1967: 69); e.g.,

/ ⁿ dimi/	‘tongues’	< {n} Noun Class 10 prefix + {limi} N stem
/ ⁿ defu/	‘tall’	< {n} Class 9 concordial prefix + {refu} Adj. stem

41.6.2.2. /n/ + /w/

The sequence /n/ + /w/ becomes prenasalized stop /^mb/ before a vowel (Polomé 1967: 69); e.g.,

/ ^m bati/	‘hut poles’	< {n} Noun Class 10 prefix + {wati} N stem
----------------------	-------------	--

41.6.2.3. *Lenition*

A large number of morphophonemic alternants result from a process of lenition whereby stops became fricatives before the derivational suffixes {i} ‘agentive’ and {y} ‘causative’ (Polomé 1967: 70–71).⁹

Examples include:

/mlifi/	‘payer’	(< {m} Noun Class 1 prefix + {lip} ‘pay’+ {i} ‘agentive’)
/mfuasi/	‘follower’	(< {m} Noun Class 1 prefix + {fuat} ‘follow’+ {i})
/mjenzi/	‘builder’	(< {m} prefix + {jenj} ‘build’+ {i})
/ruša/	‘fling’	(< {ruk} ‘fly’+ {y} ‘causative’+ {a} ‘indicative’)

This process is no longer productive in Swa., however, and stops may now be found in this context also, e.g.

/mkopi/	‘borrower’	(< {kop} ‘borrow’ + {i} ‘agentive’).
---------	------------	--------------------------------------

41.7. Historical overview

N & H (61) reconstruct the phonological system seen in Table 41-3 for Proto-Sabaki¹⁰ (also distinctive vowel length and possibly distinctive pitch-accent).

As mentioned above, all the Swa. dialects have 5-vowel systems, in which the super-close **i* and **u* have merged with /i/ and /u/ respectively. Tone and distinctive length have also been eliminated in all dialects except for Chimwini (Southern Somalia). In this section I will outline the major phonological changes affecting consonants as hypothesized by N & H.

41.7.1. *Wide Northern Dialect group* ()

The Wide Northern group includes Chimwini, Tikuu, Siyu, Pate, Amu, Mombasa, and perhaps also Chifundi and Malindi. These dialects have undergone the following changes:

9. Spirantization due to the causative suffix, reconstructed as **-y-* by N & H, is found throughout the Sabaki languages, cf. N & H 129–31. A related process is “Bantu Spirantization,” traceable even further back than Proto Northeast Coast Bantu, in which stops became fricatives, or in some cases affricates, before “super-close” **i* (see N & H 113ff.).

10. **W* is a “bilabial approximant”; **j* and **u* are the “super-close” vowels. N & H are not sure whether /r/ was a distinct phoneme in Proto-Sabaki or an allophone of /l/ (cf. p. 103); the phonemic status of /ɲ/ in Proto-Sabaki is also debatable (p. 148). N & H admit that the asymmetrical voiced consonant series **W l j g* is anomalous, but argue that it fits the facts better than a symmetrical system with all stops or all fricatives (pp. 88ff.).

Table 41-3. Proto-Sabaki Segments

Consonants			
p	t	c	k
W (b)	l (d)	j	g
f	s	š	
v	z		
mp	nt	ɲc	ŋk
mb	nd	ɲj	ŋg
rn	n	ɲ	(ŋ?)
w	(r?)	y	
Vowels			
ɨ		ʉ	
i		u	
e		o	
a			

(a) glide deletion: **fwa* > *fa*; **lʏa* > *la* (*ʏa* in Chimwini);

(b) loss of **g* in both stem-initial and intervocalic contexts.

The above changes have also affected some other dialects as well, so are not diagnostic for subgrouping (N & H 504). Northern Dialect innovations are:

(c) dentalization: **c*, **ɲc*, and **ɲj* became /t̪/, /n̪t̪/, and /ɲ̪d̪/ respectively, possibly under the influence of neighboring Cushitic languages (cf. Nurse 1985);

(d) **nk* > /h/ word-initially in certain morphemes (N & H 504).

Some additional changes affect only subsets of the northern group, outlined below.

41.7.1.1. Narrower Northern Dialect group (excluding Mombasa)

These dialects show the following changes (N & H 507):

(a) Proto-Swa. **f*i, **vi* became /si/, /zi/;

(b) Proto-Swa. **ɟ* > *ž* > *y* > *ø*;

(c) rhotacization: /^hd/ > [n̪dr], possibly in order to distinguish this phoneme from the prenasalized dental /^hd̪/;

(d) vowel assimilation in certain demonstratives: **CuCo* > *CoCo*.

41.7.1.2. "Core" Northern Dialect group (excluding Chimwini)

After Mombasa and Chimwini split off, the remaining northern dialects changed as follows (N & H 509):

- (a) **nz* > /^hd/, extending the earlier Dentalization process;
- (b) **my* > /p/;
- (c) /y/ was lost in certain morphological contexts.

41.7.2. Southern Dialects

The southern dialects (Vumba, Mtang'ata, Pemba, Tumbatu, Makunduchi, Zanzibar, Mafia, and possibly Kilwa and Mgao) share a process called "Strengthening" by N & H (cf. pp. 133ff.): some consonants are retained in contexts where they are lost or subject to lenition in other dialects and other Sabaki languages. These are usually cases where the consonant was immediately preceded by super-close **j* [i]. For example:

- **j*Wuyu 'baobab fruit' > /buyu/ in southern dialects, but /uyu/ in Giryama (Sabaki language of N. Tanzania)
- **j*lomo 'large mouth' > /domo/ in southern dialects, but /omo/ in Amu (Swa. dialect of N. Kenya)
- **j*gamba 'outer shell' > /ga^mba/ 'scale, fruit skin', but /a^mba/ in Tikuu (Swa. dialect of N. Kenya)
- **j*Wa 'steal' > /iba/, but /iwa/ in Mwani (Sabaki language of N. Mozambique)

N & H argue that Strengthening may have been caused by the fact that super-close **j* involved minimal oral aperture; they suggest that loss of **j* (e.g., in the first three examples above, where it was the prefix of Noun Class 5), where it occurred, was a later development (pp. 139–40).

As mentioned in § 41.1, the Southern Dialect group is less cohesive than the Northern groups. To a certain extent, it is defined negatively, as having not undergone the changes that have affected the Northern dialects.

Bibliography

- Ashton, Ethel O. 1944. *Swahili Grammar (including Intonation)*. London: Longmans.
- Batibo, Herman M. 1989. "Evolution and Dialectisation of Standard Kiswahili." In Rombi 1989: 149–56.
- . 1990. "Vowel Length: The Forgotten Distinctive Feature in Swahili?" *Afrikanistische Arbeitspapiere* (Schriftenreihe des Kölner Instituts für Afrikanistik, University of Cologne) 23: 53–72.

- Contini, Ellen. 1969. "Swahili Phonotactics: A Columbia School Approach." Unpublished ms.
- Engstrand, Olle, and Abdulaziz Lodhi. 1985. "On Aspiration in Swahili: Hypotheses, Field Observations and an Instrumental Analysis." *Phonetica* 42: 175–87.
- Guthrie, Malcolm. 1967–71. *Comparative Bantu*, 4 vols. Farnborough, Hant, England: Gregg International.
- Harris, Zellig. 1951. *Structural Linguistics*. Chicago: University of Chicago Press.
- Hinnebusch, Thomas. 1975. "A Reconstructed Chronology of Loss: Swahili Class 9/10." In *Proceedings of the Sixth Conference on African Linguistics* (Ohio State University Working Papers in Linguistics 20), ed. Robert K. Herbert, pp. 32–41.
- Johnson, Frederick. 1939. *A Standard Swahili–English Dictionary*. London: Oxford University Press.
- Kelly, John. 1989. "Swahili Phonological Structure: A Prosodic View." In Rombi 1989: 25–32.
- Lambert, H. E. 1958. *Chi-Chifundi, a Dialect of the Southern Kenya Coast*. Kampala: East African Swahili Committee.
- Maw, Joan, and John Kelly. 1975. *Intonation in Swahili*. London: School of Oriental and African Studies.
- Meeussen, A. E. 1967. "Bantu Grammatical Reconstructions." *Africana Linguistica* 3: 79–121.
- . 1979. "Vowel Length in Proto-Bantu." *Journal of African Languages and Linguistics* 1/1: 1–8.
- Meinhof, Carl. 1932. *Introduction to the Phonology of the Bantu Languages*, trans. and rev. in collaboration with N. J. van Warmelo and Alice Werner. Berlin: Reimer [German orig., 1910.].
- Myachina, Yekaterina N. 1981. *The Swahili Language: A Descriptive Grammar*, trans. G. L. Campbell. London: Routledge and Kegan Paul [Russian orig., 1960].
- Nurse, Derek. 1982. "The Swahili Dialects of Somalia and the Northern Kenya Coast." In *Études sur le Bantu Oriental (Comores, Tanzanie, Somalie, et Kenya)*, ed. Marie-Françoise Rombi, pp. 74–146. Paris: Société d'Études Linguistiques et Anthropologiques de France.
- . 1985. "Dentality, Areal Features, and Phonological Change in Northeastern Bantu." *Studies in African Linguistics* 16.3: 243–79.
- Nurse, Derek, and Thomas Hinnebusch. 1993. *Swahili and Sabaki: A Linguistic History*. Berkeley and Los Angeles: University of California Press.

- Nurse, Derek, and Thomas Spear. 1985. *The Swahili: Reconstructing the History and Language of an African Society, 800–1500*. Philadelphia: University of Pennsylvania Press.
- Petryankina, V. I. 1964. "O glasnykh fonemakh v yazyke suakhili" [Swahili vowel phonemes]. *Narody Azii i Afriki* 3: 119–22.
- Philippson, Gerard. 1993. "Tone (and Stress) in Sabaki." In Nurse and Hinnebusch 1993: 248–65.
- Polomé, Edgar. 1967. *Swahili Language Handbook*. Washington, D.C.: Center for Applied Linguistics.
- Port, Robert, and Kenneth Shephardson. 1982. "Morphophonemics of Swahili Verb Suffixes." *Studies in African Linguistics* 13/3: 249–71.
- Rombi, Marie-Françoise, ed. 1989. *Le Swahili et ses limites: Ambiguïté des notions reçues*. Table ronde internationale du CNRS (Sèvres, 20–22 avril 1983). Paris: Editions Recherche sur les Civilisations.
- Steere, Edward. 1870. *A Handbook of the Swahili Language Spoken at Zanzibar*. London: Society for the Promotion of Christian Knowledge.
- Tucker, Archibald N. 1946–47. "Foreign Sounds in Swahili." *Bulletin of the School of Oriental and African Studies* 11: 854–71, 12: 214–32.
- Tucker, Archibald, and E. O. Ashton. 1942. "Swahili Phonetics." *African Studies* 1: 77–103, 161–82.
- Vitale, Anthony. 1982. "Problems of Stress Placement in Swahili." *Studies in African Linguistics* 13/3: 325–30.
- Wald, Benji. 1987. "Swahili and the Bantu Languages." In *The World's Major Languages*, ed. Bernard Comrie, pp. 991–1014. London: Croom Helm.
- Whiteley, Wilfred H. 1969. *Swahili: the Rise of a National Language*. London: Methuen.

Sango Phonology

James A. Walker and William J. Samarin
University of Toronto

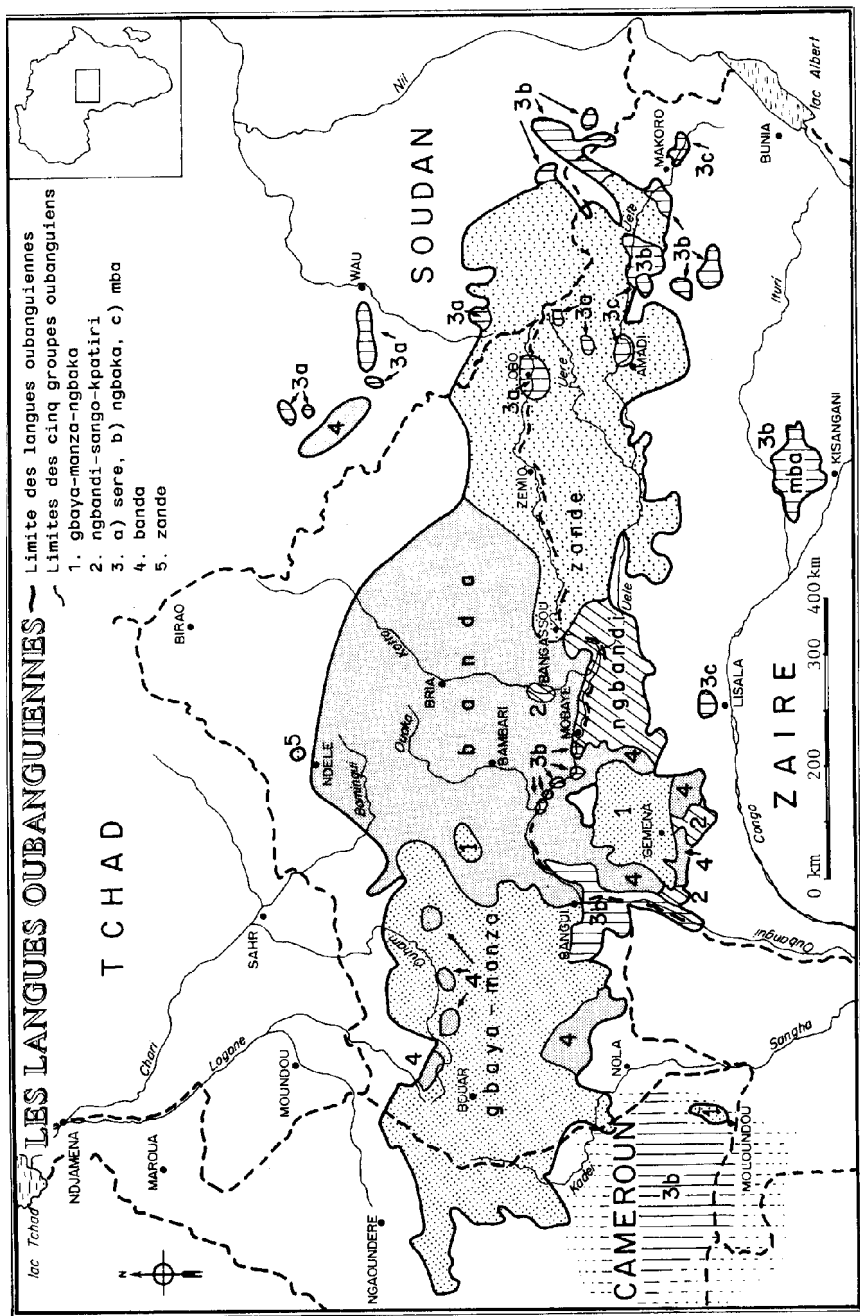
42.1. Introduction

Sango, once the national language (*langue nationale*) of the Central African Republic (CAR), but since 1991 co-official with French, is both a typical and an unusual language of central Africa. In some respects, it conforms to the phonological pattern that characterizes the region, but in other ways, it deviates from it. No doubt these deviations are due to the unusual circumstances of the origin (or formation) and development of the language: from pidgin to creole.¹

Unlike the majority of documented pidgin and creole languages, Sango is based on an African, rather than a European, language. Although it came into existence primarily as the result of colonization by Europeans in the Ubangi river basin in the last half of the nineteenth century, European languages, the most important being French, appear to have played little role in the formation of Sango. (For a discussion of the origins of the language, the reader is referred to Samarin 1982, 1989.) Apart from a number of loanwords—mostly French, a few from Portuguese, probably via Kikongo—the language was wholly African (for a discussion on French loanwords in Sango, see Taber 1964 and, more recently, Samarin and Walker 1992).

Sango is derived from a language which consists of closely related dialects called Ngbandi, Yakoma, Dendi, and Sango (see the map), which are members of the Ubangian branch of the Adamawa-Ubangian sub-family of the

1. AUTHORS' NOTE: Walker was invited by Samarin to collaborate in writing this paper, using the data collected and analyzed in connection with the project entitled "The Linguistic Consequences of the Creolization of Sango," of which Samarin is director and principal investigator. Gratitude is expressed to the Social Sciences and Humanities Research Council (Canada) for the grant that makes this project possible and to the Mission Évangélique des Frères for all its help in the field. Earlier versions of this chapter were read and commented on by Henry E. Rogers, Alan S. Kaye, and Robert Hagiwara, whose help is gratefully acknowledged. The map, used with permission, is taken from Monino 1988.



Map 5. Ubanguian Languages

Niger-Congo language family (see Samarin 1971, Boyd 1989). The pidginized form of this language, once called *sango ti turugu* 'soldier's Sango' and *sango ti gala* 'market Sango' by Central Africans and *sango véhiculaire*, Pidgin Sango (called this for the first time by Samarin in 1958), or simply Sango, was used as the medium of communication in the upper Ubangi river basin by, as one might expect, the local population, French and Belgian colonizers, including government officials, traders, and missionaries, and was partially codified by missionaries. At the independence of the Central African Republic, Sango became the national language of the country; increasingly it has become the primary, although not necessarily the first, language for most of the population of Bangui, the capital.

In this paper, we will provide an outline of the phonological system of Sango, not only in its "traditional," or pidgin, form (based on Samarin's experience with the language from 1952 to 1966, the grammar of 1967 having been based on data collected in 1962), but also in its present, evolving form (studied by Samarin with thirty weeks of fieldwork conducted in 1988, 1991, and 1992). First, we will describe the consonantal, vocalic, and tonal systems of the pidgin form of the language. Next, we will discuss our findings concerning the changes which these systems are undergoing among the younger population of Bangui. Finally, we will compare the phonological system of Sango in its present state to that of other Ubangian and central African languages. This last section will provide a background for some comments on Sango's status as a language of central Africa, and in it we will speculate about what might be expected in its phonology in the future.

Because of this temporal perspective, this chapter might well be entitled "Sango Phonology: Past, Present, and Future." It should be understood that, for reasons that will become clear, this chapter presents only some salient features of the changes that have recently been taking place.

42.2. Sango phonology: Past

In its pidgin state, Sango was described (Samarin 1967: 17) as "a demonstrably unified language, the varieties of which are few and not well defined." (For a description of Pidgin Sango, the reader is also referred to Diki-Kidiri 1977.) Although there was some variation according to the first language of the speaker, there was a predictable form to most lexemes. With some exceptions that are due largely to interference from the first language of the speaker, Pidgin Sango has a relatively simple and well-defined phonemic

inventory, and a simple (C)V(V) syllable structure.² The following sections treat of the vocalic, tonal, and consonantal aspects of Sango's phonology.

42.2.1. Vowels

Pidgin Sango has seven oral segmental vowels: /i e ε ɔ o u/. These vowels can be arranged in the following way:

Table 42-1. Pidgin Sango Vowels

	Front	Central	Back
High	i		u
Mid-tense	e		o
Mid-lax	ε		ɔ
Low		a	

However, the actual phonetic occurrence of any vocalic phoneme tends to vary (Samarin 1967: 34). For example, /i/ may occur phonetically somewhere between [e] and [i]. Furthermore, the mid-tense and mid-lax vowels in many words vary from speaker to speaker without any apparent correlation with the social characteristics of the speaker. (One would, of course, expect some correlation with the ethnic background and age of speakers.) However, slight variation in pronunciation does not affect the overall seven-vowel system. Some contrasting examples are given below:

<i>bí</i>	'night'	<i>ku</i>	'to wait'
<i>be</i>	'to weigh down'	<i>ko</i>	'to germinate'
<i>bé</i>	'liver'	<i>kɔ</i>	'to pluck'
<i>báâ</i>	'to see'		

De-syllabification, as described in Samarin 1967, is a common feature of combined vowels. This process usually occurs in front and back vowels that follow a consonant and precede another vowel. Front vowels (/i e ε/) become palatal glides and back vowels (/u o ɔ/) become labio-velar glides. Some common words in which de-syllabification occurs are:

2. The current state of Sango makes it difficult to stick to either the past or the present tense in describing the pidgin form of the language. In this section, we have used Samarin's research from the 1960s as a basis. However, Pidgin Sango is still spoken by adults, especially in rural areas. For the purposes of readability, we will stick to present-tense use for descriptions of the phonological systems, and to past tense for cited variations.

<i>kúí</i>	‘to die’	[kwí]
<i>kóé</i>	‘all’	[kwé]
<i>kàà</i>	‘hair’	[kwà]
<i>bíá</i>	‘song’	[byá]

The reasons for interpreting /w/ after a consonant as a de-syllabified vowel were primarily structural. One could argue also for a /Cw/ analysis. Samarin now writes Sango in that fashion. It should also be noted that because of contractions, /kw/ is no longer the only cluster of this kind: /tw, ndw/, possibly among others that have not yet been noted, also occur in urban Sango.

There are also nasalized counterparts of /ɛ/, /a/, /ɔ/, and /u/. The nasalized vowels /ẽ/ and /õ/ are articulated, at least by some speakers, lower than their non-nasal counterparts (Samarin 1967: 38). The phonemic nature of nasalization can be seen in the following pairs:

<i>fũ</i>	‘to smell’	<i>fú</i>	‘to sew’
<i>yõ</i>	‘to drink’	<i>yò</i>	‘to be long’
<i>hã</i>	‘to clear up’	<i>há</i>	‘to weave’
<i>kè</i>	‘to refuse’	<i>kê</i>	‘to be’ (one of several variants)

Although there are long vowels—which do not contrast with short vowels—they have a marginal status in Sango. This marginality is due to the fact that some words alternate between long and short vowels and to the fact that there are no long vowels with a level tone. (This was true of the pidgin; in urban Sango this is not true.) For example, long-vowel words like *báá* ‘to see’ and *fáá* ‘to cut’ often occur as [bá] and [fá] respectively, with the deletion of the second vowel and its accompanying tone. But even these may have been early forms of contraction. (The diachrony of contractions in Sango is dealt with in Samarin 1993b. The contractions of *tí* ‘of’ are discussed in Samarin 1992 and those of *kè* ‘to be’ in Samarin 1993a.)

42.2.2. *Tones*

Sango has three phonemic tones: high, mid, and low. High tone will here be signaled by an acute accent on the tone-bearing vowel (*á*) and low tone by a grave accent (*à*). Mid tone will be unmarked, except on syllabic nasals, when it will be marked by a macron. The three-way distinction by tone can be seen in the following examples:

<i>kà</i>	‘and, then’	<i>ka</i>	‘to sell’	<i>ká</i>	‘sore’
<i>tò</i>	‘to send’	<i>to</i>	‘to draw (water)’	<i>tó</i>	‘to throw’

Whereas three register tones were recognized by Samarin as early as 1952, because of his experience with Gbaya speakers of Sango, who tended to merge mid with high, his early publications show few mid tones.

On some words with vowel–vowel sequences, there is a change in tone from one vowel to the next, as in the following examples:

<i>báà</i>	‘to see’	[HL]
<i>zíà</i>	‘to put’	[HL]
<i>tàa</i>	‘real, true’	[HM]

In this analysis it is said that each vowel has a tone, long vowels being analyzed as a sequence of two identical vowels. Contour tones are not a feature of Pidgin Sango—that is, changes in tone occur only on long vowels (interpreted as a sequence of two identical vowels) or combinations of two dissimilar vowels, and never in a single vowel. Others would like to write /*báà*/ ‘to see’ as <*bâ*>.

Desyllabification can affect the tonal patterns. Thus, [*zíà*], just cited, becomes [*zá*].

42.2.3. *Consonants*

The consonants of Sango can be arranged as in Table 42-2.

Table 42-2. Sango Consonants

	Labial	Alveolar	Velar	Labio-velar	Glottal
Stops	p	t	k	kp	
	b	d	g	gb	
	mb	nd	ng	ngb	
	b'				
Nasals	m	n			
Fricatives	f	s			h
	v	z			
	mv	nz			
Glides		y	w		
Sonorants		l			
		r			

Some variation in the phonetic realization of consonants was noticed as correlating with ethnicity, or first language. For example, it was noted in 1962 (Samarin 1967: 31) that /*t*/ occurred as [tʃ] before /*i*/ among Banda speakers (that is, persons whose language was one of the Banda dialects). More is said about /*t*/ below.

Other, more “natural” variation, unrelated to ethnicity (that is, the first or substratal language of a speaker), also exists, such as palatalization of /s/ before /i/ (as in /sioní/ > [šioní], more frequently reduced to [šoní]). The variation of [r] and [l] (discussed further below) is also not considered typical of any one ethnic group (except for dialect differences within the Gbaya-Manza group of languages; see Samarin 1959). The glottalized stop /bʔ/ occurs in the speech of some speakers, varying with /b/, in only a few words such as /kóbè/ ‘food’.

The stop consonants /p t k kp/ are usually not aspirated, but may be aspirated depending on the first language of the speaker and the style of discourse. Today in urban speech aspiration is used, along with length and intensity, as a feature of expressivity. Some contrasts between stop consonants are given below:

<i>pa</i>	‘to slander’	<i>kà</i>	‘and, then’
<i>tà</i>	‘pot’	<i>kpà</i>	‘to resemble’
<i>dà</i>	‘house’	<i>bàà</i>	‘to see’
<i>gba</i>	‘in vain’		

Contrasts between fricatives are given below:

<i>fà</i>	‘to show’	<i>və</i>	‘to buy’
<i>sí/si</i>	‘to arrive’	<i>zi</i>	‘to dig’
<i>nzi</i>	‘to steal’		

The prenasalized bilabial fricative /mv/, occurring in only a few words, is considered marginal. The pronunciation of /z/ and /nz/ as [dʒ] and [ndʒ], respectively, is another variation associated with Banda (Samarin 1967: 33) and possibly dialects of other languages.

The fricative /h/ is often realized as a glottal stop or even deleted by some speakers. For example, /hǝ/ ‘to pass’ can occur as [ʔǝ] or [ǝ].

Although there is a great deal of variation in the pronunciation of the liquids /l/ and /r/, they are phonemic. For example, ‘child’ can occur as [mérèngé] or [mélèngé]. (Lekens noted variation between these liquids in Ngbandi [1952].) However, there is a contrast between the two phonemes in some words, as in /wala/ ‘or’ and /wara/ ‘to find’. Even words in which these consonants vary are found with one or the other predominating. The phoneme /r/ can occur as a voiced lateral flap, an apico-dental flap, or even as a trill, at least in some words, depending on the ethnic background of the speaker. The apico-dental flap seems to be preferred, and it is even replacing the lateral flap in Gbeya in Bangui, a fact noted by Samarin.

The glides /w/ and /y/ are similar to the English equivalents, except that, when the following vowel is nasalized, they may be nasalized as well, or, in

the case of /y/, may occur as a palatal nasal. For example, the word /yè/ ‘what’ often occurs as [nyè].

In summary, the phonology of Pidgin Sango can be characterized as having been and still is, when the language is used as a second language, very much like that of its parent language. It is not even a pidginized version of that original phonology—at least, as far as the inventory of phonemes and their phonetic realization is concerned.

42.3. Sango phonology: Present

In this section, we will discuss some of the phonological changes that have been discovered in urban or creolized Sango.³

The phonology of urban young people in Bangui today is more varied than it was in the pidgin state of the language. In the urban, creolized state, there seems to be more variation, rather than less. In fact, the changes are so numerous and diverse that they defy a systematic and comprehensive analysis. Despite some patterns in a few variants, the phonological form of many frequently used lexemes in young people’s speech is not predictable from a knowledge of Pidgin Sango. Many of these lexemes vary in their phonological form from one occurrence to the next, even when occurring within the same morphophonemic environment and spoken by the same speaker.

In studying the phonology of urban Sango, a sample of 130 texts was selected from a corpus of 475 texts collected by Samarin in Bangui in 1988. These texts are spontaneous discourses which were recorded and transcribed by Central African assistants trained in linguistics. Each text represents a different speaker. The topics of discourse include descriptions of pictures from a children’s book, descriptions of the preparation of meals and the performance of chores, as well as traditional tales and personal narratives. The texts were selected on the basis of the age of the speaker. Within each age group, as wide a range of education levels as possible was chosen. Ages range from 2 to 16 (although there were unfortunately very few texts for the ages of 2, 3, and 16, so that these age-groups are under-represented), and education levels range from unschooled to first-year Lycée (secondary school). Texts were also chosen to provide a variety of ethnicities, but the number of different ethnic groups available and the over-representation of

3. We would like to make it clear that in this section the language we are describing is based on extemporaneous, unmonitored speech, which is quite different from the written form of the language.

some groups (for example, Banda and Manza) in the original corpus make equal representation in the smaller corpus impossible.

42.3.1. *Phonological processes*

Although a thorough analysis of all of the changes taking place in Sango is made difficult by the interaction of many processes, it is nevertheless possible to make some generalizations. Some of the more prevalent processes involved are:

- a. vowel elision (through both deletion and vowel-shortening)
- b. deletion of intervocalic sonorants
- c. assimilation of vowels
- d. incorporation of French phonemes into Sango words

These processes are not mutually exclusive: for example, the word /méřéngé/ ('child') can occur as [mēřéngé] (deletion of intervocalic /r/) and then be further reduced to [mēřéngé] (vowel elision) or even [mēřé].

These processes will be referred to under the general label of *contractions*. It will be seen that contractions have resulted in some important changes in the language, not only in the phonology, but also in the morphophonemics and morphology of Sango, which previously had little or no morphophonemics and little morphological complexity. (For examples of common contractions, see § 42.5.)

For this analysis we have concentrated, for illustrative purposes, on words which, for grammatical reasons, have a high frequency of occurrence. Three words often involved in contractions are Sango's only two prepositions, /tí/ 'of' and /nà/ 'to, with, for, et al.',⁴ and the definite noun-phrase marker, /ní/. Other words and phonological features will also be examined where relevant to the discussion.

In contractions of these three words there are three logical possibilities: elision of the vowel, elision of the consonant, or assimilation of the vowel. Again, these three processes are not mutually exclusive. For example, there are cases in which the vowel assimilates to a neighboring vowel and the consonant is deleted. If the vowel is deleted, then the accompanying tone may be lost, or transferred to a neighboring vowel or sonorant.

The contracted forms of /ní/, /tí/, and /nà/ display all of the logical possibilities of contraction. A list of the phonological forms of the morphemes that occur in the corpus, and examples of each, are given below:

4. Samarin (1971: 129) gives figures for the frequency of /tí/ and /nà/ in texts: in one corpus, they accounted for 15.19% of all tokens, and 16.63% of all non-French tokens.

Contracted forms of ní

1. *ní* *mbò ní à-ga* 'the dog arrived' (vowel elision, tone transfer)
[from /mbò ní à-ga/]
2. *n* *mbò n-à-ga* (vowel elision without tone transfer)
3. *ná* *mbò ná-à-ga* (assimilation to following vowel)
4. *n'á* *mbò n'á à-ga* (palatalization of /n/, assimilation)
5. *nápè* *à-mú nápè* '(3 sg.) didn't take it' (vowel elision, tone transfer)
[from /à-mú ní apè/]
6. vowel *lò ga-à-á* 'he/she came with it' (consonant deletion, assimilation) [from /lò ga nà ní/]

Contracted forms of tí

1. *t-'* *mà-má t-mbí* 'my mother' (vowel elision, tone transfer) [from /mà-má tí mbi/]
2. *tC* *à-yé-t bá* '(3 sg.) wanted to see' (vowel elision without transfer) [from /à-yé tí báà/]
3. *tV* *mà-má t-álà* 'their mother' (vowel elision without transfer)
[from /mà-má tí álà/]
4. *t'V⁵* *mà-má tš-álà* ('palatalization' of /t/, vowel elision)
5. *tá* *mà-má tá-álà* (assimilation to following vowel)
6. *í* *à-yé í tè* '(3 sg.) wanted to eat' (deletion of consonant) [from /à-yé tí tè/]
7. tone *dà-á lò* 'his house' (deletion of lexeme, tone transfer) [from /dà tí lò/]

Contracted forms of nà

1. *à* *à-mú à mbi* '(3 sg.) gave (it) to me' (deletion of consonant)
[from /à-mú nà mbi/]
2. *n* *à-mú n-álà* '... to them' (vowel elision, no transfer) [from /à-mú nà álà/]
3. *là* *à-mú là kóli* '... to (a) man' (/n/ = /l/) [from /à-mú nà kóli/]
4. *V* *à-gò ò mbà* '(3 sg.) went with me' (consonant deletion, assimilation) [from /à-gòè nà mbi/]
5. *ò* *mbi zín-gò ò ndapré* 'I wake up in the morning' (vowel elision, tone transfer) [from /mbi zín-gò nà ndapéréré/]
6. *yà* *i yà mà-má* 'me and Mom'; lit. 'we with mother' (/n/ ⇒ /y/) [from /i nà mà-má/]

5. "t'" indicates a palatalized /t/, which can be realized as [tʰ] or, affricated, as [tʃ] or [ts], as in Québécois. This variant occurred in the speech of one girl in 1962. There is one Sango-Yakoma dialect with this phone before front vowels.

These contracted forms are distributed among speakers as follows:

<i>ní</i>		<i>tí</i>		<i>nà</i>	
Form	Speakers	Form	Speakers	Form	Speakers
<i>ní</i>	75	<i>t-'</i>	38	<i>à</i>	70
<i>n</i>	10	<i>tC</i>	20	<i>n</i>	8
<i>ná</i>	15	<i>tV</i>	14	<i>là</i>	8
vowel	7	<i>t'V</i>	6	vowel	2
<i>n'á</i>	2	<i>tá</i>	2	<i>ñ</i>	1
<i>nápè</i>	2	<i>í</i>	1	<i>yà</i>	1
Total	111		88		90

Despite a great deal of variation, there do seem to be certain contractions which are more “popular” than others, such as [ní], which occurs in 75 of 111 speakers who contract /ní/; [à], in 68 of 90 speakers who contract /nà/; [t-'], in 38/88 texts and [tC] (where C represents some consonant) in 20 of 88 speakers who contract /tí/. But because of the lack of a single type of contraction occurring in all texts, and the fact that different contractions of the same word can occur in the same text in the same morphophonemic environment, it is apparent that there is no consensus among the speakers.

There are other examples of contraction that have not been investigated systematically, but which should be noted:

1. Prenasalized stops becoming nasals

mbéní > *mén* ‘some, certain’

mbùnzú > *mùnzú* ‘white person’

2. Progressive assimilation of vowels

mbi ìk londo < *mbi èkè londo* ‘I get up’

3. Regressive assimilation of vowels

èpá ákpòwò < *et puis àkpòrò* ‘And then (it) boils’

4. Consonant clusters resulting from vowel deletion

srangɔ < *sarangɔ* ‘doing’ (nominalized form of *sàrà* ‘to do’)

skúlà < *sùkúlà* ‘to wash’

5. Vowel sequences resulting from intervocalic deletion of consonants.

káákɔ < *kárákɔ* ‘peanut’

kpòò < *kpòrò* ‘to boil’

42.3.2. Sociolinguistic factors

Assuming that contractions are symptomatic of language change, one would expect some correlation with age, gender, social class, and ethnicity. In only a few cases have correlations been found.

The age of the speaker, the criterion by which the texts were selected, reveals no strong correlations. The number of speakers who contracted /ní/, /tí/, and /nà/ in some way, and the average by age-group of speakers who contract (in square brackets), are given in Table 42-3 (because all contractions are considered together and some speakers use more than one kind of contraction, some figures may be higher than the number of speakers):

Table 42-3. Contractions of Particles

Age	Number of Speakers	Contractions		
		<i>ní</i>	<i>tí</i>	<i>nà</i>
2	1	0 [0]	0 [0]	3 [3]
3	2	0 [0]	0 [0]	0 [0]
5	5	2 [0.4]	0 [0]	2 [0.4]
6	7	4 [0.57]	6 [0.86]	6 [0.86]
7	7	5 [0.71]	2 [0.29]	2 [0.29]
8	15	16 [1.06]	7 [0.46]	9 [0.6]
9	12	8 [0.67]	5 [0.42]	9 [0.75]
10	15	12 [0.8]	8 [0.53]	9 [0.6]
11	15	11 [0.73]	7 [0.47]	8 [0.53]
12	16	15 [0.94]	9 [0.56]	12 [0.75]
13	16	14 [0.88]	19 [1.19]	12 [0.75]
14	10	8 [0.8]	10 [1]	8 [0.8]
15	7	9 [1.29]	11 [1.57]	7 [1]
16	2	3 [1.5]	4 [2]	3 [1.5]

The figures show an increase in the incidence of contractions in 5-year-olds and up, but the numbers remain relatively constant between the ages of 6 and 15, even when converted to a “per capita” basis by age-group. The increase from the age of 5 may have to do with the increasing complexity of the child’s discourse, the transition from “child language” to the vernacular, or perhaps his/her entry into school.

Gender, on the other hand, appears to correlate with some features. The contractions of /ní/, /tí/, and /nà/ are distributed as follows:

Contraction	Females	Males
<i>ní</i>	68 [96%]	40 [68%]
<i>tí</i>	37 [52%]	35 [59%]
<i>nà</i>	49 [69%]	40 [68%]

Although there is some variation in usage, the differences in the contractions of /*tí*/ and /*nà*/ are not significant. Females contract /*ní*/ substantially more than males do.

Even stronger correlations with gender are displayed by two other features, two phonological and one morphophonemic. These features, and the figures associated with them, are given below:

Feature	Females	Males
Palatalization/affrication	24 [34%]	8 [14%]
French [ʀ] in Sango	26 [37%]	11 [18%]
Contracted / <i>álà</i> /	13 [18%]	26 [44%]

In other words, almost three times as many females as males palatalize /*d*/ and /*t*/ before /*i*/, and almost twice as many use French [ʀ] in Sango words. Conversely, almost twice as many males as females contract /*álà*/ ('they, you (pl.)') in some way.

Social class, although problematic, reveals no correlation. An attempt was made to analyze contraction on the basis of the school the child attended, because the neighborhood (and hence the school) of the child would be determined by the socioeconomic status of his/her parents. However, there was an inadequate representation of different schools in the corpus, because the texts were not selected by this criterion. Also, a large percentage of the texts from the 1988 corpus were collected on the streets, not in the schools, and the speaker's school was not always asked for by the Central African research assistant. The results (with number of texts) for those whose school was known are as follows:

	Contracted <i>tí</i>	Contracted <i>ní</i>
Ecole Koudoukou (23)	7 [58%]	11 [92%]
Ecole Centre-Ville (12)	10 [40%]	18 [72%]

(Ecole Centre-Ville is one of the two "downtown" schools attended by children of the more affluent families of Bangui.) Despite the lack of data, it can still be seen that there is only a slight difference in incidence of contraction between the students of Ecole Koudoukou and Ecole Centre-ville. This difference does not appear to be significant.

The ethnicity of the speaker yields some unexpected correlations. Because the contractions of frequently used words displayed no strong correlation to ethnicity, three other phonological features were examined:

- i. palatalization or affrication of /d/ and /t/ before /i/
- ii. the presence of [ə] in Sango words
- iii. the use of French [ʀ] in Sango words

Whereas schwa often occurs, predictably, as a phonological variant in French loanwords, its occurrence in native Sango words is noteworthy. In the latter case, only /i ε a u/ have been observed to have schwa as a variant. Based on the occurrence of [ə] in French words in girls' dance songs (Wenezoui-Deschamps ms.), there is evidence to suggest that this variant might have been introduced, or at least been reinforced, by French—if so, this is a good example of convergence in creolization.

The last feature, [ʀ], is not a very common sound in Central Africa (reported to be in three Banda dialects far from Bangui [Cloarec-Heiss 1978]) and therefore would be expected to correlate with the ethnicity of the speaker. It comes as no surprise that this is in fact the case, although we do not know precisely what dialect of Banda was spoken by both or either of the parents.

What is surprising is the patterning of the first two features, palatalization and [ə]. Because palatalization and schwa are also found in Banda (Cloarec-Heiss 1986), one would expect them to be markers of "Banda Sango." (As was mentioned in § 42.2.1, Samarin [1967: 17] explicitly noted this fact for palatalization.) However, an examination of the incidence of these features by ethnic group does not fulfil this expectation:

Palatalization/Affrication:

Ethnic Group	Occurred in
Banda [17 texts]	4 texts [24%]
Manza [22 texts]	6 texts [27%]
Ngbaka [16 texts]	7 texts [44%]
Ngbaka-Manza [4 texts]	3 texts [75%]

[all other ethnic groups had only 1 or 2 occurrences]

Schwa in Sango words:

Ethnic Group	Occurred in
Manza [22 texts]	2 texts [9%]
Yakoma [21 texts]	4 texts [19%]
Banda [17 texts]	5 texts [29%]
Ngbaka [16 texts]	5 texts [31%]

[all other ethnic groups had occurrence in 1 text only]

As these figures show, it is in the Ngbaka-Manza (of the Gbaya language group) and Ngbaka (or Ngbaka-Ma'bo) groups, respectively, that the high-

est incidence of palatalization and [ə] occurs. This fact is particularly surprising, because grammars of the Ngbaka language (e.g. Thomas 1963) make no mention of either palatalization or [ə]. One might explain this result by saying that speakers of these ethnolinguistic backgrounds had at one time imitated an urban model with these features and surpassed it in frequency. This “hypercorrection” is similar to the “overgeneralization” of rules that occurs in pidginization (see Samarin 1966a, 1971).

The contraction /tn(e)/ (from /tènè/ ‘to speak’ or /tene/ ‘word, speech’) was also noted as a feature of “Banda Sango” in the 1960s. Because this contraction did not occur often enough to chart by ethnic group, a frequency count was performed for those texts in which /tn(e)/ did occur:

Contractions of *tènè/tene* > *tn(è/e)*:

Ethnic Group	Texts	Occurrences	Frequency
Banda	5	2	0.4
Manza	4	3	0.75
Yakoma	1	1	1
Ngbaka-Manza	1	2	2

Again, it can be seen that the occurrence is higher in the Ngbaka-Manza texts than in the Banda.

It should be noted that the incidence of all these features is still low even in the Ngbaka texts—that is, the features have not become a marker of “Ngbaka Sango.”

42.3.3. *Sango and other central African languages*

To compare Sango with other central African languages, we have looked at what has been written about the phonology of Yakoma, Ngbandi, Gbaya, Banda, and Ngbaka. These languages, apart from being substrates of the pidgin, are also adstratal, in that they continue to exert an influence on its development—many of the speakers of Sango speak one of these languages as their first or ethnic language. In terms of the phonemic inventory, Sango conforms to the central African pattern. The only consonant that it lacks is the labio-dental flap /ɱ/, which occurs in Gbeya (Samarin 1966b: 17), some dialects of Banda (Cloarec-Heiss 1978), and in Kanre (Karré), an Adamawan language. In addition to the nasalized vowels of Sango /ẽ/ and /ɔ̃/, Yakoma has nasal /ĩ/ and /ũ/, and Ngbandi has a greater number of words with nasalized vowels than Sango (Lekens 1958).

In comparison with the phonological features common to Adamawa-Ubangian languages (Boyd 1989), Pidgin Sango is a typical central African

language: it has a 7-vowel system, 3 tones, and a basically CV syllable structure. It is atypical in its lack of morphophonemic use of tone and its lack of contour tones. The lack of morphemic tone is understandable in a situation of pidginization: the morphotonemics of the various substrate languages are not identical, and they differ markedly from the tonal patterns of the Bantu languages further south. Based on a model of pidginization as “linguistic negotiation” (Thomason and Kaufman 1988) or “linguistic excavation” (Keesing 1988), it is expected that conflicting and complicated tonal systems would cancel each other out and form a “lowest common denominator” system.

Urban Sango also conforms to the 7-vowel, 3-tone system, but it further conforms in its incipient morphemic use of tone. Where it does not conform is in its unusual syllable structure, allowing more complexity than a (C)V(V) language. However, some Ubangian languages (such as Gbeya) do, with some restrictions, allow syllable-final consonants and syllable-initial consonant clusters (see, for example, Samarin 1966b), so this discrepancy is not as unusual as it might seem.

42.4. Sango phonology: Future

In this section, we review the phonological changes discussed above and attempt to demonstrate that, in terms of its use of tone, Sango is becoming more similar to some other Central African languages than it was.

In the development of Pidgin Sango into what can be called modern, “urban” Sango, some elements of the phonology have not changed. Sango still has the same basic phonemic inventory: although there is some variation in the phonetic realization of some phonemes, this variation is determined to a large extent by the morphophonemic environment and is therefore allophonic. The same can be said for the tonal inventory.

However, the phonological changes discussed above have other, more general, consequences for the language.

1. As was mentioned earlier, there is more variation, both between speakers and by the same speaker, than there was before. Because of the introduction of foreign phonemes and the processes of reduction, many lexical items no longer have a phonetic form that can be predicted from their “underlying” form.

2. Another consequence is a change in the syllable structure: because of the deletion of vowels, which results in syllable-final consonants, syllable-

initial consonant clusters, and syllabic nasal consonants, Sango has developed from a CV language into a C(C)V(C) language.

3. Finally, and perhaps most importantly, the vowel deletions and transfer of tone have led to a change in the tonal morphophonemics. The contractions associated with the attributive preposition *tí*, especially the process which we have called “*tí*-tonalization,” have altered the morphophonemic nature of Sango. In a language in which there was previously no morphophonemics, *tí*-tonalization has led to what is, in essence, a morphemic use of tone.⁶

Some of the changes from pidgin to urban Sango are not as unlikely as they may seem. Lekens (1952: ix), in describing the characteristics of Ngbandi, one of the languages from which Pidgin Sango was derived, discussed intervocalic consonant deletion and deletion of vowels in two-vowel sequences (see also Samarin 1955: 256). These observations would suggest that some of the processes of contraction are not a recent phenomenon, but are in fact merely the logical outcome of the generalization of rules of contraction that exist in the substrate.

By this last assertion, we are by no means suggesting that the substratal languages “pre-program” the pidgin for future changes, but rather that the presence of indigenous languages plays a greater role in the development of a pidgin than has previously been acknowledged. But this should not be a surprising comment in view of all that we know about changes that result from language contact.

If the process of pidginization can be viewed as the overgeneralization of rules, perhaps creolization (understood here as nativization or indigenization—i.e., acquiring native speakers) sometimes involves further overgeneralizations. After all, creolization does not take place in a linguistic vacuum, so it should not surprise us if it turns out to be less a process of *ex nihilo* creation, as some would have had us believe, than an ongoing exploitation of the linguistic resources available to speakers.

Finally, one will have noticed that some of the phonological changes that have been observed in Sango might be characterized as “natural” ones. If this is indeed what they are, they contribute to the study of language change by demonstrating that such changes can take place in a generation.

6. Samarin (1992) gives evidence that *tí*-tonalization is altering the morphology of the language. He reports having found in one text [bìík m̃] for /bik tí mbi/ (‘my pen’), an example of the possible morphologization of pronouns in possessive expressions.

Table 42-4. Urban and Rural Features

Speaker ^a	Gender	Ethnicity	Age	% ^b	Full Inventory of Urban Features in All Samples
01	F	Kanre	10	55.29	<i>m̄</i> [mid tone] < <i>mu</i> 'to take,' <i>m̄ama d</i> [or <i>n</i>] <i>lò</i> < <i>m̄ama tí lò</i> 'his mother'
02*	F	Gbaya	18	58.24	<i>Lólò ní àga</i> < <i>Lólò ní àga</i> 'Lolo (det.) came', <i>àtí nà l</i> [< <i>lì</i>] <i>ty</i> [<i>tí</i>] <i>ál</i> [< <i>álà</i>] <i>úsè</i> 'met the two of them' (lit. fell on head of 3 pl. pron. two), <i>nà á</i> [< <i>yá</i>] <i>tí ngùndà</i> 'in the bush' (lit. at belly of bush)
03	M	Mixed	12	35.71	
04	F	Pana	15	34.32	
05*	F	Mixed	11	52.12	<i>mboko lò</i> < <i>màbokò tí lò</i> 'his hand'
06*	M	Ngbugu	8	37.87	
07	F	Gbaya	7	17.26	<i>á</i> [< <i>álà</i>] <i>é byà</i> 'they sang a song'
08	M	Gbaya	12	52.66	<i>ámbo ní</i> [< <i>ní</i>] 'the dogs' (lit. dogs det.), <i>álà gwa</i> [< <i>gwè</i>] <i>àlìndà dà</i> 'they went and entered the house' (lit. they go enter house)
09*	M	Togolese	9	57.2	<i>fàsó</i> < <i>fàdesó</i> 'now,' <i>ákóli ní</i> 'the men'
10	F	Gbaya	11	13.02	<i>àgwò</i> [< <i>gwè</i>] <i>nà ngùndà</i> , <i>lì t álà ús àwè</i> '(they) went to the bush, there were two of them' (lit. go at bush head of 3 pl. pron. two finished), <i>álà gòy mbìlì</i> 'they went for a long time,' <i>lége ní áy</i> [< <i>àkè</i>] <i>úsè</i> 'there were two paths' (lit. path det. is two)
11*	F	Togolese	12	50.00	<i>fàsó</i> 'now,' <i>tèè</i> < <i>tènè</i> 'say,' <i>ángáá</i> [< <i>yanga tí</i>] <i>lò</i> 'her mouth,' <i>ángá tí lò</i> 'her mouth'
12	M	Gbaya	13	57.65	<i>tòà</i> [< <i>tòngànà</i>] <i>zò sá</i> [< <i>só</i>] <i>àmu pàpà</i> 'when someone takes the spoon' (lit. when person this takes spoon), <i>i kè báá ní</i> [< <i>í í</i>] 'we'll see'
13	M	Kaba	17	52.07	
14*	M	Mixed	10	43.37	
15*	F	Togolese	6	45.51	<i>yàyaá</i> [< <i>yàya tí</i>] <i>lò</i> 'her aunt'
16	M	Kanre	11	36.09	<i>mbi gwè nà yangá dàá</i> [< <i>yanga tí dà tí</i>] <i>bàba tí mò</i> 'I went to your father's house' (lit. I go at mouth of house of father of 2 sg. pron.), <i>fàfàtsó</i> < <i>fàdesó</i> 'now'
17*	M	Mixed	14	44.91	<i>koro</i> < <i>kotro</i> or <i>kotoro</i> 'village'
18*	F	Mbati	15	52.98	

a. * = Urban speaker.

b. Percentage of all subjects who identified this speaker as urban.

42.5. Appendix: Urban and Rural Speakers

In a study to determine whether speakers of Sango could distinguish between urban and rural varieties of the language (Samarin 1993b), 171 inhabitants of Bangui, capital of the Central African Republic, were asked to answer questions about relative age, gender, and ethnicity, in addition to provenience (urban vs. rural) of persons speaking Sango, based on what they heard in a tape recording. The results are presented in Table 42-4.

References

- Boyd, Raymond. 1989. "Adamawa-Ubangi." In *The Niger-Congo Languages*, ed. John Bendor-Samuel, pp. 178–215. Lanham, Md.: University Press of America.
- Cloarec-Heiss, France. 1978. "Etudes préliminaires à une dialectologie banda." In *Etudes comparatives*, ed. R. Boyd and F. Cloarec-Heiss. Paris: Société d'Etudes Linguistiques et Anthropologiques de France.
- . 1986. *Dynamique et équilibre d'une syntaxe: le banda-linda de Centrafrique*. Paris: Société d'Etudes Linguistiques et Anthropologiques de France.
- Diki-Kidiri, Marcel. 1977. *Le sango s'écrit aussi*. Paris: Société d'Etudes Linguistiques de France.
- Keesing, Roger M. 1988. *Melanesian Pidgin and the Oceanic Substrate*. Stanford: Stanford University Press.
- Lekens, Benjamin. 1952. *Dictionnaire Ngbandi (Ubangi-Congo belge)*. Editions de Sikkel.
- Monino, Yves, ed. 1988. *Lexique comparatif des langues oubanguiennes*. Paris: LACITO [Langues et Civilisations de Tradition Orale].
- Samarin, William J. 1955. "Sango, An African Lingua Franca." *Word* 11(2):254–67.
- . 1958. "The Phonology of Pidgin Sango." *Word* 14: 62–70.
- . 1959. "Prospecting Gbaya dialects." *African Studies* 18: 68–73.
- . 1966a. "Self-annulling Prestige Factors among Speakers of a Creole Language." In *Sociolinguistics*, ed. William Bright, 188–206. The Hague: Mouton.
- . 1966b. *The Gbeya Language: Grammar, Texts, and Vocabularies*. Berkeley & Los Angeles: University of California Press.
- . 1967. *A Grammar of Sango*. The Hague: Mouton.

- . 1971. "Adamawa-Eastern." In *Current Trends in Linguistics*, ed. Thomas A. Sebeok, vol. 7, *Linguistics in Sub-Saharan Africa*, pp. 213–44. The Hague: Mouton.
- . 1982. "Colonization and pidginization on the Ubangi river." *Journal of African Languages and Linguistics* 4: 1–42.
- . 1989. *The Black Man's Burden: African Colonial Labor on the Congo and Ubangi Rivers, 1890-1900*. Boulder, Colo.: Westview Press.
- . 1992. "Suffixation from Phonological Change in Creolized Sango." Paper presented at the 22nd Conference on African Languages and Linguistics, University of Leiden, August 30–September 1.
- . 1993a. "Morphophonemic 'Complication' in Creolization." Paper presented at the Annual Meeting of the Society for Pidgin and Creole Linguistics, Los Angeles, January.
- . 1993b. "Creolization of Pidgin Morphophonology." Paper presented at the Semi-Annual Meeting of the Society for Pidgin and Creole Linguistics, Amsterdam, 11 June.
- Samarin, William J., and James A. Walker. 1992. "Children Crafting Creolized Sango." Paper presented at Sociolinguistics Symposium 9, Reading, England, April 2.
- Taber, Charles R. 1964. "French Loan Words in Sango." M.A. thesis, Hartford Seminary Foundation, Hartford, Conn.
- Thomas, Jacqueline M. C. 1963. *Le parler Ngbaka de Bokanga*. The Hague: Mouton.
- Thomason, Sarah G., and Terence Kaufman. 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley and Los Angeles: University of California Press.
- Wenezoui-Deschamps, Martine. ms. "L'éducation sexuelle par la chanson: Les chants de fillettes en République Centrafricaine."

6 Altaic Languages

This page intentionally left blank.

Turkish Phonology

Bernard Comrie

University of Southern California

Modern Standard Turkish, together with its various dialects, is spoken as a first language by about 90% of the total population of the Republic of Turkey (estimated at 56.5 million in 1990) and as a second language by most of the rest, and as a first language by approximately one-fifth of the total population of Cyprus (estimated at 0.7 million in 1990), by Turks in other parts of the Balkans, especially Bulgaria (perhaps 9% of the total population, estimated at 9 million in 1990), and by present and former guest-workers and their dependents and descendants in north-western Europe, especially Germany (perhaps 1.5 million Turkish speakers), making a total number of native speakers of around 55 million. Turkish is by far the most thoroughly investigated Turkic language, and its phonology is one of the most thoroughly investigated of any language of the world, having had a major impact on the development of phonological theory. The main sources used for the present descriptive account are: Lees (1961), Lewis (1967), Underhill (1976).

In discussing Turkish phonology, it is useful to consider three chronological layers of vocabulary, although it should be emphasized that all three are an integral part of the contemporary language: first, items of indigenous Turkic origin; second, items of Arabic and Persian origin, here combined as Arabic–Persian (many items of Arabic origin came to Turkish via Persian, so it would be difficult in any event to separate out the two strands); third, more recent loans, primarily from Balkan and western European languages.

43.1. Segmental phonemes

The taxonomic phonemic segments of Modern Standard Turkish, other than oppositions of length in vowels (for which see below) are as shown in Tables 43-1 and 43-2. The symbols used throughout are those of Turkish orthography, but with the addition of subscript _ɰ (superscript ^ˆ in the case of *g*) to indicate front velars and the front lateral, of the macron to indicate vowel

length, and of an acute accent to indicate stress falling other than on the last vowel of the orthographic word (more than one acute per word indicates alternative stresses). (The circumflex accent, used in Turkish orthography, may be disregarded as redundant, given these other diacritics.) Phonemic transcriptions, enclosed between slants, are given only where necessary. Hyphens are used to indicate morpheme boundaries.

Table 43-1. Turkish Vowels

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i	ü	ɪ	u
Mid	e	ö		o
Low			a	

The phonetic distinction between mid and low vowels is redundant, and for phonological purposes they may be grouped together as non-high. In addition to these short vowels, Turkish also has phonemically distinct long vowels. In loans from Arabic and Persian, the long equivalents of *a*, *e*, *i*, and *u* occur, giving rise to near-minimal pairs, e.g. *sākin* ‘quiet’ (Arabic *sākin*) versus *sakın* ‘beware’, *mēm̄ur* ‘official’ (Arabic *ma’mūr*) versus *meme* ‘breast’, *īman* ‘faith’ (Arabic *īmān*) versus *imam* ‘imam’, *sūret* ‘manner’ (Arabic *šūra(t)*) versus *surat* ‘face’. Such long vowels normally occur in open syllables, though there are some examples in closed syllables, e.g. *māʔlūm* ‘known’ (Arabic *maʔlūm*). These long vowels correspond either to original long vowels in the foreign language, or to sequences of a vowel plus certain consonants (Arabic *hamza*, ‘*ayn*, sometimes *hā*). Such long vowels are not consistently reflected in Turkish orthography; here, they are indicated by a macron over the vowel in question. For all speakers, long equivalents of the back vowels *a*, *ɪ*, *o*, and *u* also occur corresponding to the orthographic sequence vowel + *ğ* before a consonant or in word final position, e.g. *dağ* /dā/ ‘mountain’, *çiğ* /çī/ ‘avalanche’, *doğru* /dōru/ ‘right’, *tuğ* /tū/ ‘banner’. Some speakers have the same phenomenon with the front vowels *e*, *i*, *ö*, and *ü*, e.g. *değ* /dē/ ‘touch’, *iğne* /īne/ ‘needle’, *öğle* /ōle/ ‘noon’, *düğme* /dūme/ ‘button’. Other speakers have a sequence of vowel + /y/ in these front-vowel examples, i.e. /dey/, /iyne/, /öyle/, /düyme/. It follows from the above that

some speakers have a complete set of long vowels corresponding to the short vowels, i.e. *ā, ē, ī, ī̄, ō, ō̄, ū, ū̄*; others have only the following long vowel phonemes: *ā, ē, ī, ī̄, ō, ū*. Long vowels are, incidentally, distinct from sequences of two identical short vowels, e.g. *dağ* /dā/ ‘mountain’, dative *dağa* /daa/; the former is one syllable, the latter two. Etymologically, the orthographic *ğ* stands for a voiced velar fricative, which was simply lost intervocalically but in syllable-final position merged with a preceding vowel to give a long vowel; in some varieties of the language, this happened only in back-vowel words, while in front-vowel words the voiced velar fricative everywhere gave /y/. This voiced velar fricative was, in native words, in complementary distribution with *g* (which latter occurred only word-initially and after an immediately preceding consonant).

Table 43-2. Turkish Consonants

	Stop	Fricative	Nasal	Other
Labial	p b	f v	m	
Dental	t d	s z	n	l ɭ r
Palatal	ç c	ş j		y
Front Velar	ķ ģ			
Back Velar	k g			
Glottal		h		

/v/ is a labio-dental fricative word-initially, a bilabial fricative between a vowel and a following consonant, and either a bilabial fricative or a frictionless continuant intervocalically; /f/ is everywhere a labio-dental fricative. /n/ has a velar allophone before a velar consonant. /ç, c/ are affricates. /r/ is usually a fricative word-initially, a tap elsewhere. The liquids /r, ɭ, l/ are often devoiced word-finally. /h/ occurs primarily in loanwords, corresponding to all three of Arabic *ḥā*’, *xā*’, and *hā*’, and some onomatopoeic items; /j/ is even more restricted to loanwords, mainly from French. In native words, the back and front velars, and likewise back (velarized) and front (non-velarized) *l* are in complementary distribution, the back variants occurring in back-vowel environments and the front variants in front-vowel environments. More specifically, the back variants occur before a back vowel, the

front variants before a front vowel; if there is no following vowel, then the front/ back quality of the preceding vowel is determinant. In loanwords, however, the front variants can occur in back-vowel environments, all three in Arabic–Persian loans, /ɫ/ also in other loans. Minimal and near-minimal pairs exist, e.g. *kar* ‘snow’ versus *ķâr* ‘work’ (Persian *kār*), *ġâvur* ‘infidel’ (Persian *gav*, variant of *gabr*), *bal* ‘honey’ versus *haɫ* ‘state’ (Arabic *ḥāl*), *roɫ* ‘role’, *vaɫs* ‘waltz’. The front velars occur only before *a* and *u*, and are indicated, albeit inconsistently, in the orthography by a circumflex on the following vowel; front *ɫ* occurs not only in this environment, and indicated orthographically in this way, but also word-finally and as part of a word-final consonant cluster, in which case there is no possible orthographic representation. In this article, front velars and front *ɫ* in a back vowel environment are indicated by a superscript or subscript hook (and in converting the representations given here to phonemic representations, the circumflex may be disregarded).

43.2. Vowel harmony

Vowel harmony in Turkish is a process whereby qualitative vowel oppositions are substantially neutralized in non-initial syllables, the quality of such a non-initial vowel being assimilated to that of the preceding vowel. Vowel harmony in Turkish proceeds from left to right through the word, although the evidence for this particular formulation, though compelling, comes more from exceptional than from regular cases.

Starting first with native words, the domain of vowel harmony is the word, including both root/stem and suffixes (there are no prefixes), except that for purposes of vowel harmony a compound word is considered as two distinct words. Within a word so defined, in non-initial syllables the only possible qualitative vowel opposition is that between high and non-high; the front/back and the unrounded/rounded oppositions are neutralized. If the first vowel of a word is front, all subsequent vowels will be front; if the first vowel is back, all subsequent vowels will be back (so-called palatal harmony). The neutralization of rounding is somewhat more complex. A high vowel will be rounded if the preceding vowel is rounded, unrounded if the preceding vowel is unrounded (so-called labial harmony); a non-high vowel in a non-initial syllable is always unrounded, i.e. there is neutralization of the unrounded/rounded distinction but no assimilation. It follows from this that the vowels *o* and *ö* occur only in initial syllables. Possible sequences of vowels are represented as follows:

<i>a, ı</i>	may be followed by	<i>a, ı</i>
<i>o, u</i>		<i>a, u</i>
<i>e, i</i>		<i>e, i</i>
<i>ö, ü</i>		<i>e, ü</i>

For certain purposes, especially assigning citation forms to suffixes, it is useful to have symbols for high and non-high archiphonemes, here represented as *I* and *A*, respectively; *I* thus stands for the appropriate member of the set *ı, u, i, ü*, and *A* for the appropriate member of the set *a, e*. Vowel harmony within stems can be illustrated by: *tarla* {tarlA} ‘field’, *ağız* {ağIz} ‘mouth’, *oda* {odA} ‘room’, *oğul* {oğIl} ‘son’, *gece* {gecA} ‘night’, *tütün* {tütIn} ‘tobacco’. For suffixes, compare the results of adding the plural suffix *-lar* and the genitive suffix *-(n)In* in the following:

<i>kadın</i>	‘woman’	<i>kadın-lar</i>	<i>kadın-ın</i>
<i>son</i>	‘end’	<i>son-lar</i>	<i>son-un</i>
<i>ev</i>	‘house’	<i>ev-ler</i>	<i>ev-in</i>
<i>gün</i>	‘day’	<i>gün-ler</i>	<i>gün-ün</i>

The left-to-right operation of vowel harmony can be seen in the fact that a word with an initial *o* or *ö* followed in the next syllable by a non-high (and therefore unrounded) vowel will be followed in the next syllable, if the vowel is high, by *ı* or *i*. Thus the genitive plural of *son* ‘end’ is *son-lar-ın*, with unrounded *ı*, as determined by the preceding *a*, not the initial *o* (which would determine a rounded vowel, as in the genitive singular *son-un*). If a syllable with *o* or *ö* is followed by a sequence of syllables with a high vowel, then this will be realized as a sequence of high rounded vowels, as in *sorum-lú-sunuz* ‘you (plural) are responsible’ (the stem is *sorumlu*, *-sInIz* is the second person plural copular clitic).

There is one fairly systematic set of exceptions to labial harmony even in native stems. In a number of stems where the vowel *a* is followed by a labial consonant, a following high vowel is rounded, as in *tavuk* ‘hen’; given the left-to-right operation of vowel harmony, the *u* of the second syllable determines the rounded variant of the high vowel in the next syllable, as in the genitive *tavuç-un*. This quirk applies, however, only within stems; across morpheme boundaries, the usual rule applies; thus the genitive singular of *çorap* ‘stocking’ is *çorab-ın*. And there are stems that do not show this phenomenon, e.g. *kapı* ‘door, gate’.

There is also a fairly systematic set of morphological exceptions to labial harmony in native stems, connected with the phenomenon of raising in verbs. Verbs whose stem ends in *A* raise this vowel to *I* before some suffixes beginning with *y*; an *I* which results from raising does not undergo labial

harmony (except before the suffix *-(I)yor*, where labial harmony does take place). There is, incidentally, a certain amount of idiolectal variation in where raising takes place, but if it does take place then vowel harmony operates as described here, as can be seen in the following examples, using the suffixes *-(y)AcAK* (future) and *-(I)yor* (progressive):

<i>anla</i>	‘understand’	<i>anlı-yacak</i>	<i>anlı-yor</i>
<i>bekle</i>	‘await’	<i>bekli-yecek</i>	<i>bekli-yor</i>
<i>kolla</i>	‘seek’	<i>kolli-yacak</i>	<i>kollu-yor</i>
<i>türe</i>	‘appear’	<i>türi-yecek</i>	<i>türü-yor</i>

In compounds, each element of the compound behaves independently with respect to vowel harmony, as in *baş-kent* ‘capital city’ (cf. *baş* ‘head’, *kent* ‘fort, town’). From the left-to-right operation of vowel harmony it follows that the last part of the compound is determining for the quality of suffixal vowels, so that the genitive singular of *baş-kent* is *baş-kent-in*, the plural *baş-kent-ler*. This principle operates even more clearly in words that are lexical exceptions to vowel harmony. A few such words are native lexical exceptions, such as *kardeş* ‘sibling’, genitive *kardeş-in*; a few native suffixes are also exceptions to vowel harmony, such as progressive *-(I)yor*, whose vowel remains *o*, therefore defining the start of a new vowel harmony domain with back vowels (and a rounded vowel, if high, in the next syllable), e.g. *gör-üyör-sunuz* ‘you (plural) are seeing’ (the stem of ‘see’ is *gör*).

Violations of vowel harmony are much more frequent, however, in loanwords, including both Arabic–Persian and other loans. Such exceptions may combine front and back vowels and may have values for rounding in non-initial syllables that violate the usual rules of vowel harmony, including *o* and *ö* in non-initial syllables. Vowel harmony continues to operate on a left-to-right basis; the exceptional vowel simply defines the onset of a new vowel-harmony domain, as in the plural and genitive singular of the following items:

<i>kitap</i> (Arabic <i>kitāb</i>)	‘book’	<i>kitap-lar</i>	<i>kitāb-ın</i>
<i>mikrop</i>	‘microbe’	<i>mikrop-lar</i>	<i>mikrob-un</i>
<i>muhip</i> (Arabic <i>muḥibb</i>)	‘friend’	<i>muhip-ler</i>	<i>muhibb-in</i>
<i>banliyö</i>	‘suburb’	<i>banliyö-ler</i>	<i>banliyö-nün</i>

Although vowel harmony most obviously affects vowels, there is also a concomitant effect on consonants. In front-vowel words, consonants are palatalized (with *k*, *g*, *l* realized as /ç/, /ğ/, /ļ/), while in back-vowel words they are not (and *l* is velarized), with the exception of *y* (which shows no variation). In words with mixed vowels, consonants are palatalized in the

same syllable as a front vowel, non-palatalized in the same syllable as a back vowel. Thus *k* is a front velar and *r* is palatalized in *kere* 'time', while *k* is a back velar and *r* is non-palatalized in *kara* 'black'. As result, suffixes also show alternation between back and front velars; thus, the lateral is back in the plural allomorph *-lar*, but front in the allomorph *-ler*. However, in some loans from Arabic and Persian, front *k*, *g*, *l* appear before the back vowels *a* and *u*, as noted in § 43.1, e.g. *kâtip* 'clerk' (Arabic *kâtib*), thus giving rise to a taxonomic phonemic opposition between these front sounds and back *k*, *g*, *l*. In addition, *l* can also appear stem-finally or as part of a stem-final consonant cluster after a back vowel in loanwords of both Arabic-Persian and more recent origin, e.g. *hal* 'state' (Arabic *ḥāl*), *rol* 'role', *kalp* 'heart' (Arabic *qalb*), *vals* 'waltz'. Such instances of final *l* define the start of a front domain for purposes of vowel harmony, so that all such words require the front variants of suffixes subject to vowel harmony, as in the plurals *hal-ler*, *rol-ler*, *vals-ler*, and the genitive singulars *hal-in*, *rol-ün*, *vals-in*. (All instances of *l* in these words are of course phonetically front.) This group of words also includes some that end in consonant + *il* in the citation form, but lack the *i* before a vowel (see § 43.3), e.g. *kavil* 'agreement' (Arabic *qawl*), plural *kavil-ler*, genitive singular *kavl-in* (front *l* throughout).

A final quirk is that certain words of Arabic (and in one case Persian) origin ending in a back vowel followed by one or two consonants idiosyncratically take front-vowel suffixes. Apart from those ending in *l* noted above (and where the final consonant is front even in the citation form), this includes certain words that in Arabic end in *kāf*, in the feminine *tā' marbūṭa* (but not the *t* of the plural ending *-āt*, or *t* resulting from final devoicing of *d*), and consonant clusters in monosyllabic stems where the final consonant is labial or dental. Examples, with plural and genitive singular, are given below. This group includes some with the front vowel *i* breaking up a consonant cluster in the citation form and before a consonant-initial suffix (in which case the stem-final consonant is phonetically palatalized), but without such a vowel before vowel-initial suffixes (see § 43.3).

<i>idrak</i> (Arabic <i>idrāk</i>)	'perception'	<i>idrak-ler</i>	<i>idrāk-in</i>
<i>saat</i> (Arabic <i>sā'a(t)</i>)	'hour'	<i>saat-ler</i>	<i>saat-i</i>
<i>harf</i> (Arabic <i>ḥarf</i>)	'letter'	<i>harf-ler</i>	<i>harf-in</i>
<i>vakit</i> (Arabic <i>waqt</i>)	'time'	<i>vakit-ler</i>	<i>vakt-in</i>

In the first three examples, the final consonant is not palatalized, i.e. there is no way of telling from the pronunciation of the citation form that they take front-vowel endings.

For at least some speakers, some words that ended in *qāf* in Arabic may take back-vowel suffixes, even though they have syllable-final front vowels in Turkish, but nowadays front-vowel suffixes are also possible, e.g. *fevk* ‘top’ (Arabic *fawq*), genitive *fevkın* or *fevkin*. The *k* is back in *fevkın*, front in *fevk* and *fevkin*. Turkish, unlike some other Turkic languages (including Tatar and Uyghur) does not have back velars (or *ʎ*) in front-vowel environments. Words like *hakikat* ‘truth’ (Arabic *haqīqa(t)*) are not exceptions, since the pronunciation is /hakiykat/ or /haki:kat/.

43.3. Syllable structure and boundary phenomena

The basic syllable structure of native words is (C)V(C)(C). A single consonant between vowels is (with possible exceptions noted below) assigned to the following syllable; two consonants between vowels are split between the two syllables; three consonants are assigned two to the first syllable, one to the second.

In syllable-initial position, there is normally one and only one consonant. Word-initially, however, vowel-initial syllables are common, e.g. *ana* ‘mother’. In native words, medial vowel-initial syllables did not originally occur, thus excluding in particular the possibility of word-internal vowel sequences. However, the loss of the voiced velar fricative has given rise to a number of word-internal vowel sequences, both within morphemes, e.g. *ağaç* /aaç/ ‘tree’, *ağır* /air/ ‘heavy’, and across morpheme boundaries, e.g. *dağ-a* /daa/, *dağ-in* /dain/, dative and genitive respectively of *dağ* /dā/ ‘mountain’. Vowel sequences are common in words of foreign origin, including words of Arabic origin, with zero reflexes of Arabic *hamza* and ‘*ayn*, e.g. *teessüf* ‘regret’ (Arabic *ta’assuf*), *saat* ‘hour’ (Arabic *sā’a(t)*). For some speakers, the sequence of consonant + *hamza*/**ayn* in an Arabic word has a reflex in Turkish pronunciation in that the consonant is assigned to the first syllable and the second syllable is vowel-initial, i.e. for such speakers *sanat* ‘art’ (Arabic *šan’a(t)*) is syllabified /san\$at/; other speakers do not make this syllabification distinction, pronouncing /sa\$nat/.

Initial consonant clusters are confined to recent loanwords; earlier loans added a prothetic or epenthetic vowel, e.g. *istasyon* ‘station’ (from French *station*), *kuruş* ‘piastre’ (from German *Groschen*); more recent loans often have alternate forms, e.g. *klüp* or (more colloquially) *kulüp* ‘club’ (from French *club*, as evidenced by the front vowel). Indeed, in native words a number of single consonants either do not occur word-initially or are heavily restricted, namely *c* (except in some onomatopoeic words), *f*, *l*, *m*, *n* (except

in the interrogative pronoun *ne* 'what'), *ı*, *z*. (Initial *m* occurs in the interrogative particle *mI*, but although this is written as a separate word, the fact that its vowel is subject to vowel harmony shows that it is phonologically part of the preceding word.) In some old loans, and more widely in colloquial usage, a prothetic vowel is inserted, e.g. *oruç* 'fast (during Ramadan)' (from (Early Modern) Persian *rōza*), *limon* (colloquially also *ilimon*) 'lemon'.

Syllable-finally, Turkish words of native origin tolerate only a very narrow range of consonant clusters, basically only those in which the first consonant is a nasal or liquid, and the second consonant a stop, e.g. *sanç* 'pierce', *kalk* 'rise', *dört* 'four'. Loans have increased this set somewhat, in particular by allowing a fricative in first or second position in the cluster, e.g. *ders* 'lesson' (Arabic *dars*), *harf* 'letter' (Arabic *ḥarf*), *şevk* 'desire' (Arabic *şawq*). However, it is still the case that most final consonant clusters in loanwords are not imported into Turkish in syllable-final position, although they may reappear when a vowel-initial suffix is added so that the syllable division falls between the two consonants. This applies to all word-final geminate consonants, which are simplified syllable-finally, as in *hat* 'line' (Arabic *xatt*), plural *hat-lar*; genitive singular *hatt-in*. Many other consonant clusters receive the epenthetic vowel *i* syllable-finally, e.g. *şehir* 'city' (from Persian *šahr*), plural *şehir-ler*, genitive singular *şehir-in*. Syllable-final clusters of three consonants are rare even in the most recent loans, the only example usually cited being *ravnt* 'round (e.g. in boxing)', from English *round*.

Word-finally, the voiced stops (*b*, *c*, *d*, *g*) are basically excluded in Turkish, in both native and loan words. However, at least for some speakers, there are some exceptions, including *ad* 'name' (which if pronounced with a final voiced consonant forms a minimal pair with *at* 'horse'), *hac* 'pilgrimage to Mecca' (Arabic *ḥajj*) (likewise contrasting with *haç* 'crucifix', a loan from Armenian).

Partly in response to the constraints on syllable structure noted above, and partly because of other constraints, Turkish shows a fair amount of morphophonemic alternation at morpheme boundaries. Some of the major alternations are discussed below.

A number of words end in a voiceless stop when this occurs syllable-finally, but a voiced stop (or its reflex, in the case of the velars) when this occurs followed by a vowel, as in the following examples:

<i>dip</i>	'bottom'	<i>dip-ler</i>	<i>dib-in</i>
<i>ağaç</i>	'tree'	<i>ağaç-lar</i>	<i>ağac-in</i>
<i>dāmat</i> (Persian <i>dāmād</i>)	'son-in-law'	<i>dāmat-lar</i>	<i>dāmad-in</i>
<i>renk</i> (Persian <i>rang</i>)	'color'	<i>renk-ler</i>	<i>reng-in</i>

This list also includes verbs, e.g. *git* 'go', aorist *gid-er*. Given that a fair number of words have a stem-final voiceless consonant in both environments, e.g. *top* 'ball', genitive *top-un*, *sepet* 'basket', genitive *sepet-in*, *saç* 'hair', genitive *saç-in*, *dok* 'dock', genitive *dok-un*, the usual analysis is that those words that show the alternation have morphophonemically a stem-final voiced consonant that is devoiced syllable-finally, while those that do not show the alternation simply have morphophonemically a stem-final voiceless consonant, i.e. {dib} but {top}. Items like *ad* 'name', which for some speakers have a voiced final consonant even word- and syllable-finally, would then need to be treated as exceptional, a kink in the analysis that does not occur in more famous examples of word- or syllable-final devoicing, such as Russian and German. (Note also that fricatives are not affected by word- or syllable-final devoicing in Turkish.) In addition, a number of rules of thumb can be given for determining whether a word shows the alternation or not, e.g. monosyllables are less likely to show it than are polysyllables, although these rules do have exceptions. Finally, it should be noted that some loanwords ending in a voiceless consonant in the donor language do, or at least may, show the alternation in Turkish, e.g. *grup* 'group', genitive singular *grub-un* (or *grup-un*).

In the case of word-final velars, the situation is slightly more complicated. Words ending in a final consonant plus velar, all of which are loans, have *k* syllable-finally and *g* syllable-initially. In native words and most loans, after a vowel one finds *k* syllable-finally and zero (corresponding to orthographic *ğ*) when there is a following vowel. Some recent loans retain syllable-final *g* in spelling and standard pronunciation (but have a nonstandard pronunciation with *k*), and either retain *g* or have zero (orthographic *ğ*) before a vowel. Examples are given below:

<i>renk</i>	'color'	<i>renk-ler</i>	<i>reng-in</i>
<i>ekmek</i>	'bread'	<i>ekmek-ler</i>	<i>ekmeğ-in</i>
<i>elektrik</i>	'electricity'		<i>elektriğ-in</i>
<i>jeolog</i>	'geologist'	<i>jeolog-lar</i>	<i>jeolog-un/jeoloğ-un</i>

Following on the analysis given above for the voiced–voiceless alternation in general, one could argue that these words end morphophonemically in {g}, which is devoiced to *k* syllable-finally, deleted intervocally, and remains between a consonant and following vowel. Words like *jeolog* would (not implausibly, given the variation in their treatment) have to be treated as exceptions.

In discussing long vowels in words of Arabic and Persian origin in § 43.1, it was noted that such long vowels normally (though not quite exclusively)

occur only in open syllables. Many such loans show alternation between a long vowel in an open syllable and a short vowel in a closed syllable, as in the following examples:

<i>zaman</i> (Arabic <i>zamān</i>)	‘time’	<i>zaman-lar</i>	<i>zamān-in</i>
<i>tēsir</i> (Arabic <i>ta’θīr</i>)	‘effect’	<i>tēsir-ler</i>	<i>tēsīr-in</i>

Given the vast number of nouns ending in VC that do not show a long vowel before a following vowel, the obvious analysis is to assume that items like those displayed above have a morphophonemic long vowel in their final syllable, which is shortened in a closed syllable, i.e. {*zamān*}, {*tēsīr*}. Words like *mālūm* ‘known’ (Arabic *ma’lūm*), with a long vowel in a closed syllable, would then need to be treated as exceptions.

Loanwords may also show alternation between a single consonant syllable-finally and a geminate consonant before a vowel, as in the following:

<i>hat</i> (Arabic <i>xaṭṭ</i>)	‘line’	<i>hat-lar</i>	<i>hatt-in</i>
<i>af</i> (Arabic <i>‘afw</i>)	‘pardon’	<i>af-lar</i>	<i>aff-in</i>

In the second example, Arabic word-final *fw* has been reinterpreted as *ff*. Morphophonemically, such words can be taken to end in a geminate consonant, which is simplified syllable-finally, i.e. {*hatt*}, {*aff*}.

The alternations noted above can be combined, as in the following examples, although there are no examples for combining shortening and degemination (reflecting the fact that in Arabic long vowels do not generally occur before a consonant cluster):

<i>kitap</i> (Arabic <i>kitāb</i>)	‘book’	<i>kitap-lar</i>	<i>kitāb-in</i>
			(shortening, devoicing)
<i>muhip</i> (Arabic <i>muḥibb</i>)	‘friend’	<i>muhip-ler</i>	<i>muhibb-in</i>
			(degemination, devoicing)

In native words, the loss of the voiced velar fricative has given rise to alternations between a long vowel (syllable-finally) and a short vowel (before a following vowel), as in *dağ* /dā/ ‘mountain’, plural *dağ-lar* /dālar/, genitive singular *dağ-in* /daɪn/.

Constraints on final consonant clusters (other than geminate consonants) give rise to alternations between sequences of stem-final CJC in syllable-final position with CC before a vowel, as in the following examples, which include both native and loan words:

<i>boyun</i>	‘neck’	<i>boyun-lar</i>	<i>boyn-un</i>
<i>isim</i> (Arabic <i>ism</i>)	‘name’	<i>isim-ler</i>	<i>ism-in</i>

Since many words end in CJC and show no such alternation, e.g. *koyun* ‘sheep’, genitive *koyun-un* (which forms a minimal pair with *koyun*, genitive *koyun-un* ‘bosom’), *deniz* ‘sea’, genitive *deniz-in*, and since one can make a

rather general statement of which consonant clusters are permitted syllable-finally in Turkish, instances of this alternation can be analyzed by assuming a morphophonemic final consonant cluster and a rule inserting a high vowel to break up this cluster syllable-finally, i.e. {boyn}, {ism}. The need for lexical specification is minimal, though there are some anomalies, e.g. *ebeveyn* 'parents' (Arabic *'abawayn*), genitive *ebeveyn-in*, shows precisely the final *yn* cluster that is avoided in *boyn* 'neck', genitive *boyn-un*.

As noted in § 43.2, words whose last stem-vowel is back but which nonetheless take front-vowel suffixes, if they take an epenthetic vowel require a front vowel, as in *kavil* 'agreement' (Arabic *qawl*), genitive *kavl-in*, *vakit* 'time' (Arabic *waqt*), genitive *vakt-in*. In such items, then, the domain of vowel harmony that embraces the suffixes also includes the final consonant and the immediately preceding position that comes to be occupied by the epenthetic vowel.

A few more oddities concern examples where, for purposes of vowel epenthesis, an original Turkish voiced velar fricative or an original Arabic consonant (*hamza* or *'ayn*) that has no direct phonetic realization in Turkish is nonetheless treated as if present. In the native Turkish examples, where orthographic *ğ* marks the position of the 'quiescent' consonant, one ends up with an alternation between stem-final *VIC* and *VC*. In the Arabic words, if one marks the position of the quiescent consonant by #, one gets, in the case of stem-final {VC#}, alternation between *VCI* and *VC*, for {VCI#} and {VC#} respectively, while for stem-final {V#C} one gets alternation between *VIC* (for {V#IC}) and *VC* (for {V#C}). These possibilities are illustrated below; note that in the loans from Arabic an apostrophe is sometimes used to indicate the quiescent consonant.

<i>oğul</i> /oul/	'son'	<i>oğul-lar</i> /oullar/	<i>oğl-un</i> /ölun/
<i>nevi</i> (Arabic <i>naw'</i>)	'sort'	<i>nevi-ler</i>	<i>nev'-in</i>
<i>kair</i> (Arabic <i>qa'r</i>)	'(sea) bottom'	<i>kair-lar</i>	<i>ka'r-in</i>

(The apostrophe in *nev'-in* may, for some speakers, represent assignment of the intervocalic consonant to the preceding syllable.) The underlying forms of the last two items are {nev#}, giving {nevi#}, i.e. /nevi/, before a consonant and {nev#}, i.e. /nev/, before a vowel, and {ka#r}, giving {ka#ır}, i.e. /kair/, before a consonant and {ka#r}, i.e. /kar/, before a vowel.

A further phonotactic constraint of Turkish is that sequences of obstruents must agree in voicing. A number of suffixes begin with a voiced stop when occurring after a vowel or voiced consonant, but with a voiceless stop when occurring after a voiceless consonant, e.g. locative *-dA*. These

suffixes can be assumed to begin morphophonemically with a voiced stop (as otherwise there would be no general rule to predict their voiced nature after a vowel or sonorant consonant), which is devoiced by a preceding voiceless obstruent, e.g. *gece* 'night', locative *gece-de*, *el* 'hand', *el-de*, but *top* 'ball', *top-ta*. Note in particular that after a morphophonemic voiced stop, as in {dib} 'bottom', one finds syllable-final devoicing of the obstruent preceding the morpheme boundary and the voiceless variant of the suffix-initial vowel, i.e. *dip-te*.

A number of suffixes have alternative forms, one beginning with a vowel and one beginning with a consonant, the former used after a preceding consonant, the latter after a preceding vowel. For suffixes of this kind with an initial variable vowel, this is almost invariably *I*, so that the first person singular possessive suffix, for instance, is *-(I)m*; adding it to *el* 'hand' gives *el-im*, while adding it to *gemi* 'ship' gives *gemi-m*. The major exception is the aorist suffix *-(I)r*, which for a number of underived monosyllabic-stem verbs ending in a consonant has the form *-Ar* rather than *-Ir*, although the latter is found with some such verbs, as well as with all polysyllabic and derived verbs, e.g. *anla* 'understand', *anla-r*, *gel* 'come', *gel-ir*, *konus* 'speak', *konus-ur*, but *dön* 'turn', *dön-er*. (This variation in the vowel of the aorist goes back to Old Turkic and is found throughout the language family.) For suffixes with a variable initial consonant, this is usually *y*, as in the first person singular copula suffix *-(y)Im*, attached to *Türk* 'Turk' to give *Türk-üm* 'I am a Turk' and to *ev-de* 'at home' (locative of *ev*) to give *ev-dé-yim*. There are, however, some idiosyncratic exceptions, such as genitive *-(n)In* and third person singular possessive *-(s)I*; added to *ev* 'house' and *tarla* 'field' these give: *ev-in*, *ev-i*, but *tarla-nın*, *tarla-sı*. While in many cases these variable vowels and consonants avoid unacceptable (or at least originally unacceptable) vowel and consonant clusters, such as **Türk-m*, **tarla-ı*, in many instances there is no such compelling motivation; thus, the first person plural possessive suffix is *-(I)mIz*, which added to *el-le*, plural of *el* 'hand', gives *el-ler-imiz*, although **el-ler-miz* would be perfectly phonotactic in Turkish (cf. *gör-mek*, infinitive of *gör* 'see').

Some words that end in a vowel nonetheless behave as if they ended in a consonant. This includes, first, words that end orthographically in *ğ* (and phonetically in a long vowel), so that *dağ* /dā/ 'mountain' has genitive singular *dağ-ın* /daɪn/. For some speakers, the Turkish derivatives of Arabic words ending in *V + hamza/ayn* take suffixes as if the stem ended in a consonant, while other speakers treat such words like all other words ending in a vowel;

thus *cāmi* ‘mosque’ (Arabic *jāmiʿ*) has genitive *cāmi-nin* or *cāmi-in*, *mevzū* ‘topic’ (Arabic *mawḍūʿ*) has genitive *mevzū-nun* or *mevzū-un*. (Note that in this last example the stem-final vowel is long, even for those speakers who treat it as {mevzū#}, with a quiescent consonant ‘closing’ the final syllable.)

43.4. Stress

In regularly stressed, non-compound words, main stress falls on the last syllable. In Turkish, this main stress has pitch as its prime phonetic concomitant, e.g. *çocúk* ‘child’. (Regular stress is indicated here by an acute accent only where it is the topic of discussion.) The addition of a suffix (as opposed to a clitic, see below) leads to movement of the stress further to the right to the last syllable of the word, e.g. plural *çocuk-lár*, first person plural possessive *çocuk-lar-ımíz*, dative *çocuk-lar-ımız-á*. In this respect, as with regard to stress in general, words of Arabic–Persian origin behave like native words, e.g. *kitáp* ‘book’ (Arabic *kitāb*), *kitap-lár*, *kitap-lar-ımíz*, *kitap-lar-ımız-á*.

A number of non-compound words have irregular stress; the following are the major classes:

- a. Place names never have final stress; most have initial stress, e.g. *Ánkara*, *Páris*, while some have other non-initial stress, e.g. *İstánbul*. Place names and common nouns that are otherwise homophonous typically differ in stress, e.g. *ordú* ‘army’, but *Órdu*.
- b. Loanwords (other than traditional loans from Arabic–Persian) often have lexically specified irregular stress, usually corresponding to the stress in the donor language, e.g. *lokánda* ‘restaurant’ (Italian *locanda*), *táksi* ‘taxi’.
- c. A number of other nouns, mainly denoting kin and animals, have lexically irregular stress, e.g. *ábla* ‘elder sister’, *hála* ‘paternal aunt’ (Arabic *xāla*), *karínca* ‘ant’.
- d. Many adverbs have initial stress, whether of native or Arabic–Persian origin, e.g. *sónra* ‘afterwards’, *évveĵá* ‘firstly’ (Arabic *ʾawwalan*), *yázın* ‘in summer’. Such adverbs may thus differ in stress from otherwise homophonous words of other word classes, e.g. *gerçek-tén*, ablative of *gerçek* ‘true, truth’, *gérçek-ten* ‘truly’, *nihāyét* (Arabic *nihāya(t)*) ‘end’, *níhāyet* ‘finally’.
- e. Some polysyllabic suffixes have stress on the first syllable of the suffix, e.g. the converb (gerund) suffix *-(y)ÁrAk*, as in *yap-árac* ‘by doing’.

- f. Interjections have initial stress, e.g. *háydi* 'come on'; this applies also to vocatives, which may thus contrast in stress with the noun in other uses, e.g. *babá* 'father', but *bába!* 'father!' In Turkish, however, it does not apply to imperatives, which are stressed regularly, e.g. singular *başlá* 'begin!'

Addition of suffixes to a word with non-final stress does not change the position of the stress, e.g. *Ánkara*, locative *Ánkara-da*. (The apostrophe here is purely orthographic, used to divide a proper name stem from a suffix.)

Compounds have their main accent on the first component, thus usually (barring the occurrence of a lexically irregular item as first component) on the last syllable of the first component, with secondary stress on the second component. This applies not only to compounds consisting of a combination of bare lexical morphemes, e.g. *báš-bakàn* 'prime minister' (cf. *baş* 'head', *bakan* 'minister'), *bír-takím* 'several' (cf. *bir* 'one', *takım* 'set'), but also to the so-called *izafet* group, in which the dependent noun takes a possessive suffix, e.g. *ye-mék oda-sí* 'dining room' (cf. *ye-mek* 'eat (infinitive)', *oda* 'room', third person possessive *-(s)í*). Under compounds are included intensive adjectives with initial reduplication, the initial reduplication counting as the first element of a compound, e.g. *çırıl-çıplàk* 'stark naked' (cf. *çıplak* 'naked'). When suffixes are added to compounds, the primary stress does not move, e.g. genitive *báš-bakan-ín*.

In addition to suffixes, Turkish also has clitics, often not distinguished terminologically from suffixes in the literature. As far as stress is concerned, clitics do not form part of the same phonological word as the preceding morphemes, and are thus either unstressed or (if forming a longer chain of syllables) have a secondary final stress. The clearest cases of clitics are those that do nonetheless undergo vowel harmony, i.e. by one criterion (stress) do not form part of the preceding word, but by another (vowel harmony) do form part of the preceding word. Such harmonic clitics include:

- a. Instrumental *-IA* 'with', e.g. *memnūniyét-le* 'with pleasure'.
- b. Adverbial *-cA*, e.g. *iyí-ce* 'well' (cf. *iyi* 'good').
- c. The negative suffix *-mA*, e.g. *anlá-ma-dí* 'he did not understand'. There is, however, one exception, namely the aorist form of verbs, where the negative suffix is stressed, e.g. *bíl-mé-z-sín* 'you do not know', literally 'know-negative-aorist-you' (for clitic *-sIn*, see (d) below).
- d. Cliticized forms of the copula, e.g. *yorgún-um* 'I am tired'. It should be noted that most verb forms are historically non-finite forms with cliticized forms of the copula, e.g.

gör-úr-sünüz (second person plural of the aorist of *gör* 'see'), though some are not and therefore have stress on the last syllable, e.g. the corresponding past tense *gör-dü-nüz*.

- e. The predicative suffix *-dIr*, e.g. *yorgún-dur-lâr* 'they are tired' (literally 'tired-predicative-plural'), cf. *yorgun-lâr* without the predicative suffix.
- f. The interrogative particle *mI*, which is written as a separate word, e.g. *gör-dü-nüz mü* 'did you (plural) see'.

(Lees (1961: 43, 48) observes that some speakers may shift the stress to the syllable immediately preceding a clitic, and gives this as the only possibility before the conjunctive particle *ki* 'so', e.g. *gél-me-di* 'he didn't come', but *gèl-me-dí ki* 'he didn't come, so ...'.)

For items that do not undergo vowel harmony (non-harmonic clitics), there is less reason to assume that one is dealing with a single word; this is the case, for instance, with postpositions, e.g. *kadín için* 'for the woman', literally 'woman for'. Some morphemes occupy an intermediate position between suffixes and clitics: thus plural *-lAr* is basically a suffix; however, in the third person plural of the negative of the aorist, it may either be stressed as a suffix (e.g. *bil-me-z-lér* 'they do not know'), or unstressed as a clitic (*bil-mé-z-ler*). Progressive *-(I)yor* is particularly idiosyncratic, also from the viewpoint of vowel harmony, since the *I* undergoes vowel harmony but the *o* does not; stress either falls on the *o* of *yor*, or (as if the *yor* part alone were a clitic) on the preceding vowel, whether this is the variable vowel of the suffix, as in *bil-íyór-um* 'I know', or the stem-final vowel, as in *anlí-yór-um* 'I understand'.

References

- Lees, Robert B. 1961. *The Phonology of Modern Standard Turkish*. Bloomington: Indiana University Press.
- Lewis, G. L. 1967. *Turkish Grammar*. Oxford: Oxford University Press.
- Underhill, Robert. 1976. *Turkish Grammar*. Cambridge: MIT Press.

Tatar (Volga Tatar, Kazan Tatar) Phonology

Bernard Comrie

University of Southern California

The longer names for plain Tatar are to avoid confusion with other Turkic languages and ethnic groups also referred to as Tatar: some of these, such as Baraba Tatar, are genetically close to Tatar within Turkic, while others, such as Crimean Tatar, are not. In this chapter, the appellation Tatar will be used in its narrow sense. Tatar is spoken by over 5 million people, nearly all in Russia; the greatest concentrations are in the Tatar republic (Tatarstan, capital: Kazan) and the Bashkir republic: in the Tatar republic Tatar and Russian have about equal numbers of native speakers, while in the Bashkir republic Tatar speakers are outnumbered by Russian speakers but in turn outnumber speakers of Bashkir, a language closely related to Tatar; Russian administrative divisions (*oblasti*) with predominant Russian population but with more than about 100,000 Tatar speakers are, in roughly descending order of Tatar-speaking population: Chelyabinsk, Ulyanovsk (Simbirsk), Orenburg, Yekaterinburg (Sverdlovsk), Perm, Tyumen. According to Soviet census statistics, in 1979 about 80% of native speakers of Tatar spoke Russian fluently as a second language. Tatars constitute, incidentally, the largest minority both ethnically and linguistically in the Russian Federation.

The standard language is based on the Central dialects, which include those of the area around Kazan. At present Tatar is written using a variant of the Cyrillic alphabet, which provides a close, but by no means exact, representation of the taxonomic phonemes of the standard language. In this chapter, a taxonomic phonemic representation will be used; stress, however, will be indicated by an acute accent only where it does not fall on the last syllable of the word. This chapter is based exclusively on secondary sources, in particular Kurbatov et al. (1969), supplemented by Poppe (1963); in Kurbatov et al. (1969), the chapter on Phonetics was written by L. T. Maxmutova, the chapters on Orthoepy and on Graphics and orthography by X. R. Kurbatov.

Three major chronological layers can be identified in Tatar vocabulary: first, indigenous words of Turkic origin; second, words of Arabic–Persian origin—for most purposes, words of Arabic and of Persian origin can be

treated together—; third, words of Russian origin, including recent internationalisms, which have been borrowed through Russian. Given the centuries during which Tatar has been under increasing Russian influence, the language contains a large number of Russian loans. While early Russian loans were assimilated to indigenous Tatar phonology, more recent loans have not been; the orthoepic norm recommends pronunciations indistinguishable from those used in Russian, although in practice one would assume that assimilation takes place to differing degrees depending on such factors as an individual's proficiency in Russian. The phonology of Russian loans in Tatar will be noted herein only in passing. By contrast, the earlier layer of loans from Arabic–Persian will be treated.

44.1. Segmental Phonemes

The taxonomic phonemic segments of Standard Tatar, excluding phonemic segments that occur only in loans from Russian, are shown in Tables 44-1 and 44-2, although the text of this section should be read for justification of and qualifications to this claim.

Table 44-1. Tatar Vowels

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i	ü	ï	u
Mid	e	ö	ə	o
Low	ä		a	

The vowel *a* is phonetically somewhat rounded in the initial syllable of a word (including the initial syllable of a non-initial component of a compound word), while occurrences of *a* in subsequent syllables have successively lesser degrees of rounding culminating in complete lack of rounding. The same applies to the vowels *o* and *ö*; in non-initial syllables the current orthography writes these with the symbols for *ə* and *e*, respectively, though here we use *o* and *ö*, as this is both more accurate phonetically and facilitates presentation of the facts of vowel harmony.

The mid (phonetically: high-mid) vowels are shorter than the others and are highly centralized; indeed they are often described as central vowels

with some degree of fronting (for *e*, *ö*) or backing (for *ə*, *o*). The phonemic status of *ĩ* is controversial; some sources using a Latin-based transcription use the symbol *ĩ* for our *ə*, and transcribe our *ĩ* as a diphthong, *ĩy*, which would correspond to *əy* in our transcription. While it is true that *ĩ* is diphthongized in some environments—syllable-finally, especially word-finally—the same is true of *i*, and since the two vowels do form a front-back pair for purposes of vowel harmony, there seems to be no synchronic reason to treat them differently. Although the main thrust of this chapter is synchronic, those familiar with other Turkic languages should bear in mind that Tatar has undergone a switch of earlier high and mid vowels in most environments: thus Tatar *e*, *ö*, *ə*, *o* correspond to *i*, *ü*, *ĩ*, *u* in most Turkic languages, while *i*, *ü*, *u* correspond to *e*, *ö*, *o*. (Tatar *ĩ* is often of secondary origin, other Turkic languages lacking a mid back unrounded vowel.)

In loans from Russian, vowels have values close to those found in Russian; in particular we find low back unrounded *a*, low-mid back rounded *o*, low-mid front unrounded *e*, high central unrounded *ĩ*.

In general, Tatar lacks phonemic distinctions of vowel length. However, the following sequences of a high vowel followed by a glide may be realized either as such or as a phonetically long variant of the vowel: *iy*, *ĩy*, *uw*, *üw*. They are essentially restricted to certain morphological forms, e.g. first person singular present tense *qarĩ-y-m* (from *qara* ‘look’), *ešli-y-m* (from *ešlä* ‘work’), verbal noun *uku-w* (from *ukə* ‘read’), *biyü-w* (from *biye* ‘dance’). (Note that in these verbal forms the final vowel of the stem undergoes raising. Here and elsewhere, hyphens are used to indicate morpheme boundaries.)

The consonants *f* (which is bilabial), *ž*, *x*, and *h* occur primarily in words of Arabic-Persian or onomatopoetic origin—*ž* more specifically in those of Persian origin; occasional exceptions in native words include *tufraq* ‘earth’ (Old Turkic *topraq*), *xatən* ‘woman’ (Old Turkic *qatun*). Tatar *x* corresponds to both Arabic *ḥ* and *x*, e.g. *xäl* ‘state’ (Arabic *ḥāl*), *xaləq* ‘people’ (Arabic *xalq*). Phonemic *ʔ* and the phonemic contrasts between front-velar *k* and back-velar *q*, between front-velar *g* and back-velar *ɣ* occur only in words of Arabic-Persian origin. The phonemic glottal stop can be seen in examples like *mäsʔälä* ‘problem’ (Arabic *masʔala*), *täʔsir* ‘influence’ (Arabic *taʔīr*). At the beginning of a breath group, initial vowels are preceded by a non-phonemic glottal stop, e.g. *at* ‘horse’ is [ʔat] in isolation; in the middle of a breath group, the glottal stop is not pronounced; if the preceding word ends in a consonant, this consonant is syllabified as the initial consonant of the following syllable, i.e. *yul östö* ‘on the way’ (‘way on’) is syllabified

Table 44-2. Tatar Consonants

	Stop	Fricative	Nasal	Other
Labial	p b	f	m	w
Dental/Alveolar	t d	s z	n	l r
Alveolo-palatal		ś ź		y
Palato-Alveolar		š ž		
Front Velar	k g		ŋ	
Back Velar	q	x ɣ		
Glottal	ʔ	h		

[yʊ\$lös\$stö]; if the preceding word ends in a vowel, then, depending on the particular combination, either a glide is inserted or the two vowels are merged into a single syllable, e.g. *yaŋa ikmäk* ‘new bread’ is syllabified [ya\$ŋa\$yik\$mäk], *qáysə orsoq* ‘which spindle’ is syllabified [qay\$sor\$soq], *qáyda uraq* ‘where is the sickle?’ (literally ‘where sickle’) is syllabified [qay\$daw\$raq]. In native words, *k* and *g* occur in the environment of front vowels, *q* and *ɣ* in the environment of back vowels. In loanwords, however, this correlation is frequently violated; see § 44.2. (It should be noted that *ɣ* is usually a fricative, though in free variation with a back velar stop, functioning in either case as the back counterpart of plosive *g*.) The nasal *ŋ* has front and back allophones depending on the vocalic environment, but these never contrast phonemically. *w* is bilabial, in word-initial position a voiced fricative, elsewhere a voiced glide. In the standard language, *ś* and *ź*, corresponding to *č* and *ž* in most other Turkic languages, are fricatives, not affricates, distinguished from *š* and *ž* by a more forward articulation; most transliterations of Tatar retain the symbols *č* and *ž*.

In loans from Russian other consonants occur, in particular labio-dental *f* and *v*, affricates *c* and *č*, the long fricative *šš*, plus the full battery of Russian palatalized consonants.

44.2. Vowel harmony

Tatar has palatal harmony and, though in a very restricted form, labial harmony. For purposes of vowel harmony, a compound word counts as two

words, i.e. vowel harmony operates independently in each component of the compound. The principles of vowel harmony are close to exceptionless in native words, though loanwords, whether Arabic–Persian or Russian, provide many exceptions. By vowel harmony, the oppositions of frontness and rounding are neutralized in non-initial syllables. By palatal harmony, if the first vowel of a word is front, then all subsequent vowels must be front, and if the first vowel of a word is back, then all subsequent vowels must be back.

In general, in Tatar non-initial vowels are unrounded, with two provisos, one general, the other idiosyncratic. If the initial vowel of a word is *o* or *ö*, then following mid vowels are rounded *o* and *ö*, respectively, rather than *ə* or *e*, although the degree of rounding of *o* and *ö* decreases progressively through the word. Note that only *o* and *ö* have this effect; after the other rounded vowels, namely *u* and *ü*, one finds *ə* and *e*, not *o* and *ö*. The idiosyncratic exception is the verbal noun suffix *-(E)w* (for this notation, see below and § 44.3), which obeys palatal harmony but often ends up phonetically (see below) with a high rounded vowel. Vowel harmony is seen most clearly when suffixes are attached. (Tatar, like Turkic in general, lacks clear instances of prefixes, in particular prefixes that would undergo vowel harmony with the stem.) Affixes have two (or four, where labial harmony is also involved) forms, depending on the front/back (and, where relevant, rounded/unrounded) quality of the preceding vowel, such as plural *-lar/-lär* (morphophonemically: *-IAr*), second person plural possessive *-(ə)Ƴəz/-(e)gez/-(o)Ƴoz/-(ö)göz* (morphophonemically: *-(E)GEz*). The full list of back–front pairs of vowels found under palatal harmony is as follows:

Back	Front	Morphophoneme
<i>a</i>	<i>ä</i>	<i>A</i>
<i>ə</i> (rounded: <i>o</i>)	<i>e</i> (rounded: <i>ö</i>)	<i>E</i>
<i>i</i>	<i>i</i>	<i>I</i>
<i>u</i>	<i>ü</i>	<i>U</i>

Note in particular that the front-vowel correspondent of *a* is always *ä*, never *i* (< Proto-Turkic *e*). Alternations involving the high vowels are very restricted. The *i/ï* alternation is found when attaching the present gerund suffix *-y* to stems ending in a low vowel: the vowel is raised to *I*, thus giving *qarï-y* from *qara* ‘look’, *ešli-y* from *ešlä* ‘work’. The *u/ü* alternation is found when attaching the verbal noun suffix *-(E)w* to a stem ending in a consonant or in a mid vowel; if the stem ends in a consonant, the variable mid vowel *E* is present, and this vowel or the final mid vowel of the stem is raised and rounded to *U*, as in *yaz-uw* (from *yaz* ‘write’), *uqu-w* (from *uqə* ‘read’), *kil-üw* (from *kil* ‘come’), *biyü-w* (from *biye* ‘dance’).

Attachment of the suffixes *-lar* and *-(E)GEz* to stems of various vocalisms is shown below:

<i>baš</i>	'head'	<i>baš-lar</i>	<i>baš-əɣəz</i>
<i>əm</i>	'gesture'	<i>əm-lar</i>	<i>əm-əɣəz</i>
<i>qul</i>	'arm'	<i>qul-lar</i>	<i>qul-əɣəz</i>
<i>qoş</i>	'bird'	<i>qoş-lar</i>	<i>qoş-oyoʒ</i>
<i>qolar</i>	'tool'	<i>qolar-lar</i>	<i>qolar-əɣəz</i>
<i>it</i>	'meat'	<i>it-lär</i>	<i>it-egez</i>
<i>tel</i>	'tongue'	<i>tel-lär</i>	<i>tel-egez</i>
<i>küz</i>	'eye'	<i>küz-lär</i>	<i>küz-egez</i>
<i>öy</i>	'house'	<i>öy-lär</i>	<i>öy-ögöz</i>

The example *qolar-əɣəz* illustrates the left-to-right nature of Tatar vowel harmony: the vowel of the third syllable is ə, conditioned by the immediately preceding a, even though the initial vowel is o; o and ö can only give rise to rounding in an immediately following syllable, although, as illustrated by *qoş-oyoʒ* and *öy-ögöz*, this rounding can be passed iteratively to the next mid vowel to the right.

The left-to-right nature of Tatar vowel harmony becomes clearer in looking at compound words whose components are of different vowel harmony sets, and simple words, almost all loans, that violate vowel harmony. In a compound word, it is always the last component that determines the front/back value of suffixal vowels. In an exceptional word, it is always the vowel of the last syllable that determines the front/back value of suffixal vowels. These possibilities are illustrated below:

<i>bil-baw</i>	'belt'	<i>bil-baw-lar</i>	<i>bil-baw-əɣəz</i>
<i>aq-qoş</i>	'swan'	<i>aq-qoş-lar</i>	<i>aq-qoş-oyoʒ</i>
<i>kitap</i> (Arabic <i>kitāb</i>)	'book'	<i>kitap-lar</i>	<i>kitab-əɣəz</i>
<i>bāha</i> (Persian <i>bahā</i>)	'price'	<i>bāha-lar</i>	<i>bāha-ɣəz</i>
<i>surāt</i> (Arabic <i>şūra(t)</i>)	'shape'	<i>surāt-lär</i>	<i>surāt-egez</i>

In the first example, *bil* means 'waist', *baw* means 'rope'; in the second, *aq* means 'white', *qoş* means 'bird'. With loans of Russian origin, however, that contain both back and front vowels and do not end in a palatalized consonant (see below), there is a convention of treating them as back-vowel words, i.e. of treating the vowels e and i as being neutral with respect to vowel harmony; thus, *sovet* 'Soviet', but *Sovet-lar* *Sojuz-ə* 'Soviet Union' ('Soviet-plural union-its'), *sovétnik* 'counselor', but *ilselek sovétnik-ə* 'embassy counselor' ('embassy counselor-its').

While palatal harmony is noticeable primarily in the vowels, it also affects the velar consonants (*k/q*, *g/ɣ*, *ŋ*), which, in native words, have back variants

in back-vowel words (e.g. *qoş* ‘bird’), but front variants in front-vowel words (e.g. *küz* ‘eye’); more specifically, the back variants occur in a syllable with a back vowel, the front variants in a syllable with a front vowel. From this it follows that *k* and *q*, *g* and *ɣ*, alternate in suffixes, as in dative *-qa/-kä*, second person plural possessive *-əɣəz/ -egez*; these consonants will be represented morphophonemically as *K* and *G* respectively. In words of Arabic–Persian origin, however, it is common for the back variants to occur in front-vowel environments, with *q* corresponding to Arabic *qāf*, as in *qälām* ‘pen’ (Arabic *qalam*), *diqqät* ‘attention’ (Arabic *diqqa(t)*), and *ɣ* corresponding to Arabic ‘ayn and *ɣayn*, as in *mälyün* ‘cursed’ (Arabic *mal’ün*), *şayır* ‘poet’ (Arabic *şā’ir*), *käyāz* ‘paper’ (Persian *kāyaz*). There are even occasional native words that show this exceptional behavior, e.g. *qärdäş* ‘kin’ (cf. Turkish *kardeş*; Tatar has accommodated the word to vowel harmony by fronting the first vowel, but without modifying the initial consonant, a good indication of the relative independence of velar and vowel quality in contemporary Tatar). The front/back value of the nasal consonant *ŋ* is completely determined by the phonetic environment.

Unlike Turkish, Tatar does not generally have front velars in back-vowel environments. In part because of its richer vowel system, with the distinct phoneme *ä*, Tatar can compensate for front/back velars by using front/back vowels, as in the initial consonant of *kägäz* ‘paper’, cited above, cf. Turkish *kâğıt* (with a front velar followed by a back vowel). There are, however, a few exceptions, e.g. *pak* ‘pure’ (with a final front velar, from Persian *pāk*). Interestingly, the word-final front velar causes this word to take front vowel suffixes, cf. the derived noun *pak-lek* ‘cleanness, purity’, as do word-final palatalized consonants in Russian loans, e.g. *sekretar* ‘secretary’, plural *sek-retar-lär*. Otherwise, Tatar does not have words with a back vowel in the final syllable but taking front-vowel suffixes. Finally, in Russian loans, the velars *k* and *g* in back-vowel environments are supposed to be pronounced as mid velars, i.e. not the back velars that one would find in this environment in an indigenous Tatar word; however, such words still take back-vowel harmony, so that the third person possessive of *barak* ‘barracks’, for instance, is *barag-ə*.

44.3. Syllable structure and boundary phenomena

The basic syllable structure of Tatar is (C)V(C)(C), and this applies to both indigenous words and words of Arabic–Persian origin; words of Russian origin often show the much more complex syllable structures of that language,

with initial consonant clusters and final clusters of more than two consonants. Vowel-initial syllables occur only word-initially, and since they are introduced phonetically by a glottal stop, one could even simplify the range of options to CV(C)(C); the vowel *ɪ* (or, equivalently, the sequence *ʔi*) is not found word-initially. The velar nasal *ŋ* is not found word-initially; the liquids *ɾ*, *l* are not found word-initially in indigenous words, but are common in loans; word-initial *d* and *z* are rare in indigenous words, but there are exceptions, including such common words as *di* ‘say’, *dürt* ‘four’. Word-finally, the voiced stops *b*, *d*, *g* (and its back variant *ɣ*) are not found; this applies to both indigenous and loan words; word-final *z* is found only in loan words, e.g. *ixtiyaž* ‘need’ (Arabic *iḥtiyāj*), *taž* ‘crown’ (Arabic *tāǧ*). Syllable-final consonant clusters are very rare in Tatar, in both indigenous words and words of Arabic–Persian origin, essentially confined to sequences of liquid, nasal, or *y* followed by a stop (or fricative *ś*, for earlier **č*), e.g. *šāns* ‘pierce’, *dürt* ‘four’, *äyt* ‘say’. However, many words that end in a consonant cluster in other Turkic languages have an epenthetic vowel in Tatar, e.g. *qərəq* ‘forty’ (cf. Turkish *kırk*), *qaləq* ‘rise’ (cf. Turkish *kalk*). Even loans from Arabic and Persian are usually made to conform to this pattern (see below).

In Tatar, the opposition between voiced and voiceless stops is neutralized in word-final position in favor of the voiceless member of the pair. For the purposes of this phenomenon, the voiced back velar fricative behaves as if it were the corresponding stop. However, the etymological affricates *ś* and *ž* do not participate in this neutralization; word-final *ž* does occur, though only in words of Arabic–Persian origin, as noted above. As in other Turkic languages, this gives rise to alternations between voiceless stops (syllable-finally) and voiced stops (before a vowel). However, while in some Turkic languages, such as Turkish (see chapter 43), there is good evidence for distinguishing morphophonemically between stem-final voiced and voiceless stops, this is not the case in Tatar. With stem-final dental stops, Tatar always has the voiceless variant, i.e. one never finds alternations between voiceless and voiced dental stops; thus *at* ‘horse; name’, third person singular possessive *at-ə*. Likewise, words ending in *ś* (for earlier **č*) show no voicing alternation, e.g. *kiś* ‘evening’, *kiś-e*. With stem-final labial stops, voicing always takes place, i.e., there are no stems that end in *p* and retain voiceless *p* prevocally; thus *tap* ‘find’, present gerund *tab-a*, and likewise in words of Arabic–Persian and even of Russian origin, e.g. *kitap* ‘book’ (Arabic *kitāb*), *kitab-ə*, *klup* ‘club’, *klub-ə*, *prīncip* ‘principle’, *prīncīb-ə*; note that of the last two words cited, both loans from Russian, the first ends morphophonemi-

cally in *b* in Russian, but the second ends in *p* in Russian and still undergoes voicing in Tatar. With stem-final velars the situation is a little more complex. In words where the stem-final velar is consistently preceded by a vowel, voicing always takes place, as in *ayaq* 'foot', *ayaɣ-ə*, *jiläk* 'berry', *jiläg-e*, *aq* 'flow', present gerund *ay-a*, *barak* 'barracks' (from Russian), *barag-ə*; in nominals of native origin where the penultimate vowel is variable (see below), voicing likewise takes place, e.g. *baləq* 'fish', *baly-ə*; however, in verbs of native origin and in words of Arabic–Persian origin, this voicing does not take place, e.g. *selek* 'wave', present gerund *selk-ä*, *xaləq* 'people' (Arabic *xalq*), third person singular possessive *xalq-ə*. It should be noted that the difference between *baly-ə* on the one hand and *selk-ä* and *xalq-ə* on the other corresponds to the difference noted below between those instances where a mid vowel may be retained even prevocally (and is retained orthographically) and those where such a vowel cannot be retained (and is not represented orthographically), i.e., the voicing takes place if and only if the velar is potentially preceded by a vowel. Voicing also fails to take place, predictably, in Russian loans that end in a consonant cluster with final velar even in citation form, e.g. *cirk* 'circus', *cirk-ə*.

The fact that the voicing takes place only before a vowel, and not before a consonant, can be seen, for instance, by adding the locative suffix *-dA* to such nouns, e.g. *at-ta*, *kitap-ta*, *baləq-ta*, *xaləq-ta*, in which the stem-final consonant is devoiced and indeed occasions devoicing of the initial consonant of the suffix.

A number of Tatar stems end in the sequence *CEC* syllable-finally, including word-finally, but drop the final mid vowel when occurring before a vowel, e.g. *awəz* 'mouth', third person singular possessive *awz-ə* 'his mouth', *iren* 'lip', *irn-e* 'his lip', *boron* 'nose', *born-o* 'his nose', *sigez* 'eight', *sigz-enşe* 'eighth', *qərəq* 'forty', *qəry-ənsə* 'fortieth', *qurəq* 'fear', present gerund *qurq-a*, *selek* 'wave', *selk-ä*. This group also includes words of Arabic–Persian origin, e.g. *waqət* 'time' (Arabic *waqt*), *waqt-ə* 'his time', *xaləq* 'people' (Arabic *xalq*), *xalq-ə*, *isem* 'name' (Arabic *ism*), *ism-e*. As far as I am aware, in Tatar, unlike in Turkish (q.v.), there is no contrast between words that retain this vowel before a vowel and those that lose it; thus, there is no reason for assuming two sets of morphophonemic representations, one for words with a stable vowel and one for words with a variable vowel. There is, however, a distinction statable in a combination of phonetic and etymological criteria. In verbs and words of Arabic–Persian origin with a final velar, the vowel is never pronounced when the stem occurs

prevocally and is not represented orthographically; the orthography thus has (in transliteration) *kurka*, *selkä*, *xalkə*. However, in all native nominals, and also in words of Arabic–Persian origin where the final consonant is not velar, variants with different degrees of presence versus absence of the vowel are permitted when the stem is prevocalic, both syllabifications being permitted, for instance, in verse; here, the orthography consistently represents the vowel, i.e. (in transliteration) *awəzə*, *borənə*, *sigezenśe*, *kəragənśə*, *wakətə*, *iseme*.

Tatar does not tolerate clusters of obstruents that do not agree in voicing, and as a result a number of suffixes beginning with *d* or *g/y* have variants beginning with *t* and *k/q*, respectively, used after stems ending in a voiceless obstruent. Thus, the dative suffix is *-dA*, as when added to *awəl* ‘village’ to give *awəl-da*, but shows up as *-tA* when added, for instance, to *at* ‘horse’ to give *at-ta*; compare likewise the dative suffix *-GA* in *awəl-ga*, but *at-qa*. Stems ending in *z* and occurring before an affix beginning with a voiceless obstruent undergo the reverse assimilation, the stem-final *z* becoming *s* under the influence of the following voiceless obstruent; compare *yaz* ‘spring (season)’, but *yas-səz* ‘springless’, *yas-śa* ‘in a springlike manner’.

Another general assimilatory process is the assimilation in place of articulation of *n* to a following bilabial stop or velar stop (including *ɣ*), as in *mīn* ‘I’, but *mīm-me* ‘I?’ (‘I-interrogative’), *un* ‘ten’, but *úm-biś* ‘fifteen’ (‘ten-five’), *yan* ‘burn’, past participle *yaŋ-yan* ‘one who has burnt’.

More restricted assimilations are the following. The initial *l* of the plural suffix *-lAɣ*, and the initial *d* of the ablative suffix *-dAn*, assimilate to give *n* after a preceding nasal, as in the following forms of *uram* ‘street’: plural *uram-naɣ*, ablative *uram-nan*. The morphologization of this rule can be seen by contrasting locative *uram-da*, where the assimilation does not take place; compare also the lack of assimilation with the derivational suffix *-lEK*, as in *utən* ‘wood’, *utən-ləq* ‘woodshed’.

In part in order to maintain the basic syllable structure of Tatar, with its avoidance of vowel hiatus and of clusters of more than two consonants, several suffixes have both vowel-initial and consonant-initial variants, the former used after stems ending in a consonant, the latter after stems ending in a vowel. The most common such pattern is for a suffix to begin basically with a consonant, but to take an extra initial vowel *E* when used after a consonant-final stem. This can be illustrated with the first person singular and second person plural possessive suffixes, *-(E)m* and *-(E)GEz* respectively, attached to the following nouns:

<i>apa</i>	‘aunt’	<i>apa-m</i>	<i>apa-γəz</i>	<i>apa-sə</i>
<i>taraq</i>	‘arm’	<i>taraγ-əm</i>	<i>taraγ-əγəz</i>	<i>taraγ-ə</i>

An anomaly in Tatar, as in Turkic generally, is that a number of consonant-final verbs in the future (aorist) take the suffix *-Ar*, rather than the regular *-(E)r*; contrast regular *kil* ‘come’, *kil-er*, with anomalous *köl* ‘laugh’, *köl-är*. The distribution of *-Ar* versus *-Er* after a consonant is subject to dialectal variation in Tatar; thus *yaz* ‘write’ forms *yaz-ar* in the standard language, while the colloquial language also allows *yaz-ər*. Less commonly, a suffix will begin basically with a vowel, but will require an extra initial consonant after a vowel-final stem; an example is the third person possessive suffix *-(s)E*, as shown in the rightmost column of the examples just presented, or the future participle suffix *-(y)AśaK*, e.g. *kil* ‘come’, *kil-äśäk*, *eślä* ‘work’, *eślä-yäśäk*. An idiosyncratic pattern is presented by the present gerund suffix, which has the variant *-A* after a consonant, but *-y* (with raising of preceding *a/ä* to *i/i*) after a vowel, e.g. *yaz* ‘write’, *yaz-a*, *qara* ‘look’, *qarī-y*. While in some instances the variation between consonant-initial and vowel-initial suffix forms has a clear phonotactic motivation—**apa-ə* and **taraq-m* or **taraγ-m* would be impossible words in Tatar—in others the alternation has simply been morphologized; there is, thus, nothing non-phonotactic about **taraq-qəz*, cf. dative *taraq-qa*. More idiosyncratic morphologization can be seen in such phenomena as the behavior of the dative suffix *-GA*, which drops its initial *G* only after possessive suffixes of the first and second person singular, e.g. *at* ‘horse’, dative *at-qa*, *at-əm* ‘my horse’, dative *at-əm-a*, *at-əbəz* ‘our horse’, dative *at-əbəz-ga*.

44.4. Stress

Regular stress in Tatar falls on the final syllable of the word. As suffixes are added, the stress shifts to the right to the last syllable, e.g. *balá* ‘child’, plural *bala-lár*, third person possessive *bala-lar-ə*, genitive *bala-lar-ə-nəŋ*. Compounds are stressed on the last syllable of the first component, e.g. *áq-qoş* ‘swan’ (‘white-bird’), *andá-monda* ‘here and there’ (‘there-here’), *bík-yaxşə* ‘very good’ (‘very-good’); this also includes affective words with reduplication (often with some modification) of the initial syllable, the reduplicated part functioning as if first part of a compound, e.g. *típ-tigez* ‘dead even’, cf. *tigez* ‘even, equal’.

Certain words are lexical or morphological exceptions to this general rule, having stress on the initial syllable. They include the following:

- a. Second-person imperative forms of verbs, e.g. singular *útər*, plural *útər-əyəz* 'sit!'.
- b. Interrogative pronouns and adverbs, e.g. *qáysə* 'which', *nísä* 'how many'.
- c. Adverbs and conjunctions, indigenous or of Arabic–Persian origin, e.g. *yáŋa* 'recently, just now', *tāmám* 'completely' (Arabic *tamām(an)*), *lākin* 'but' (Arabic *lākin*).

This phenomenon can give rise to minimal pairs, contrast the adjectives *yaŋá* 'new' and *tāmám* 'complete' (Arabic *tamām*).

In addition to suffixes, there are also clitics, which are unstressed themselves and leave stress on the preceding sequence. Clitics include:

- a. The interrogative clitic *-mE*, e.g. *bar-də-lár-mə* 'did they go?', cf. *bar-də-lár* 'they went' ('go-past-plural');
- b. The negative clitic *-mA*, e.g. *bár-ma-də-lar* 'they did not go'. Note that in the positive the stress is *bar-də-lár*, i.e., the negative clitic not only cannot take the stress itself but also prevents the stress moving further to the right. There is, however, one exception: the negative behaves as a suffix, taking stress regularly, in the future (aorist), e.g. *ešlā-má-s* 'he will not work', *ešlā-mä-béz* 'we will not work'.
- c. The item *dA* 'also', conventionally written as a separate word, although it undergoes vowel harmony, as in *sín dā* 'you also'.
- d. The adverbial clitic *-sA*, e.g. *tatár-sa* 'in Tatar'.
- e. The clitic *-dAy* 'like', e.g. *sín-däy* 'like you'.
- f. All predicative clitics, e.g. *yaz-ár-mən* 'I shall write' ('write-future-first:person:singular'), *yaz-ár-səz* 'you (plural) will write', *mín šayír-men* 'I am a poet', *sín qazan-nán-səŋ* 'you are from Kazan' ('you Kazan-ablative-second:person:singular'). As in Turkic languages generally, many verbal forms are historically non-finite forms with predicative clitics, as in the first two forms just listed; contrast a finite verb form like past tense *qara-də-m* 'I worked', *qara-də-yəz* 'you (plural) worked'.

In addition one might include here various unstressed items that do not undergo vowel harmony, e.g. postpositions, as in *eš ösön* 'for the sake of work', literally 'work for'. The difference between suffix and clitic can give rise to minimal pairs distinguished only by stress, e.g. *almá* 'apple', *ál-ma* 'do not take!', *sín-dä*, locative of *sín* 'you', *sín dā* 'you also'. An intermediate position between suffix and clitic is occupied by the plural suffix *-IAr*. Usually,

it is a suffix, as in *bala-lár* 'children', *bar-də-lár* 'they went'; but in the third person plural of the present tense, it is a clitic, e.g. *yaz-á-lar* 'they write'.

The above applies to indigenous words and words of Arabic–Persian origin. Russian loans, especially recent ones, usually retain stress on the same syllable as in the original, and the addition of inflectional suffixes does not change this stress, e.g. *fábrika* 'factory', plural *fábrika-lar*.

References

- Kurbatov, X. R.; L. T. Maxmutova; L. P. Smoljakova; and È. R. Tenišev, eds. 1969. *Sovremennyj tatarskij literaturnyj jazyk: Leksikologija, fonetika, morfologija*. Moscow: Izd-vo Nauka.
- Poppe, Nicholas. 1963. *Tatar Manual* (Indiana University Publications, Uralic and Altaic Series, 25.) Bloomington: Indiana University; The Hague: Mouton.

This page intentionally left blank.

Uyghur Phonology

Bernard Comrie

University of Southern California

Uyghur (other spellings: Uygur, Uighur, Uigur) or Modern Uyghur, a member of the Eastern (Chaghatay) branch of the Turkic family, is spoken by some seven and a half million speakers (Grimes 1992: 529), mainly in the Xinjiang Uyghur Autonomous Region in northwestern China (over seven million), though including also some 200,000 speakers in the Central Asian republics of Kazakhstan and Kyrgyzstan. The variety of Uyghur treated is the modern standard language, based on the Northern (also called: Central) dialects. The following is based on the excellent description and analysis provided by Hahn (1991).

The Uyghur lexicon can be divided into three main chronological layers, and this distinction is relevant to the phonology of the language. First, there are indigenous words of Turkic origin. Second, there are words borrowed from Arabic and Persian; for most purposes, these do not need to be distinguished, and this layer of the lexicon can be referred to as Arabic–Persian. Finally, there are words borrowed from Russian and Chinese (Mandarin). At present Uyghur is written in either Arabic script (in China) or in Cyrillic script (in the former Soviet Central Asian republics); until recently in China, Roman script was also used. In this chapter, a taxonomic phonemic transcription is used, except that stress is marked only where it is irregular or the point at issue.

45.1. Segmental phonemes

The qualitatively distinct taxonomic vowel phonemes of Uyghur are presented in Table 45-1.

The absence of a high unrounded back vowel is striking in comparison with most other Turkic languages, reflecting a historical merger of earlier **i* and **ĩ*. Morphophonemically, it is still necessary to distinguish between “front *i*” and “back *i*” in Uyghur, as will be seen in § 45.2, as some instances of this single taxonomic phoneme behave as a front vowel, others as a back

Table 45-1. Uyghur Vowels

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i	ü		u
Mid	e	ö		o
Low	ä		a	

vowel. Back *ɨ*, incidentally, never occurs word-initially. (A phonetic *ɨ* is possible in some styles of pronouncing Russian loanwords in Uyghur, corresponding to the Russian high back/ central unrounded vowel.) In indigenous and Arabic–Persian words, the vowel *e* occurs almost exclusively as a morphophonemic alternant of *a* or *ä*; it is, however, frequent in Chinese and Russian loans. Some occurrences of *e* in loanwords behave as a front vowel morphophonemically, others as a back vowel, so that here again it is necessary to distinguish between “front *e*” and “back *e*”—and note that even in indigenous and Arabic–Persian words the *e* that alternates with *ä* can be considered front, the *e* that alternates with *a* as back. At the morphophonemic level, then, the vowel system of Uyghur can be presented as in Table 45-2, with the vowels in parentheses occurring only in loanwords.

Table 45-2. Uyghur Morphophonemic Vowel Qualities

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i	ü	ɨ	u
Nonhigh	ä, (e)	ö	a, (ə)	o

High vowels are devoiced when unstressed between voiceless consonants (one of which may be the word-initial glottal stop).

Uyghur has in addition a phonemic opposition between short and long vowels, mainly in the Arabic–Persian and Chinese–Russian layers of its vocabulary. Some instances of vowel length directly reflect Arabic–Persian length, e.g. *tästīq* ‘ratification’ (Arabic *taṣḍīq*), *hawā* ‘air’ (Arabic *hawā*), and some reflect reinterpretation of Chinese and Russian phonology, e.g. *pilān*

‘plan’ (Russian *plan*), *dūfū* ‘tofu, beancurd’ (Chinese *dòufu*). There are minimal pairs distinguished only by length, e.g. *tār* ‘sweat’, *tār̄* ‘complexion’. The phonetic distinction between long and short vowels is often neutralized when the vowels are unstressed, although, as noted in § 45.2, short and long vowels continue to behave differently morphophonemically. In addition, long vowels also occur phonetically as alternative pronunciations of sequences of short vowel plus liquid (*ɿ*, *ʌ*), sometimes also *y*, word-finally or before a consonant, as in *bar* (also [bā]) ‘go’, *kāl-gān* (also [kāl̄gān]), past participle of *kāl* ‘come’, *hoyla* (also [hōla]) ‘courtyard’. Such phonetically long vowels resulting from absorption of a liquid or *y*, however, do not behave morphophonemically other than as short vowels.

The taxonomic consonant phonemes of Uyghur are presented in Table 45-3, excluding some that are found only in Chinese–Russian loans.

Table 45-3. Uyghur Consonants

	Stop	Fricative	Nasal	Other
Labial	p b		m	w
Dental	t d	s z	n	l r
Palatal	č ǰ	š ʝ		y
Front Velar	k g		ŋ	
Back Velar	q	x ɣ		
Glottal	ʔ	h		

The phoneme *b* is normally a bilabial fricative between vowels or between a vowel and a liquid. The consonants *č* and *ǰ* are normally affricates, but are often pronounced as (and merge with) the corresponding fricatives syllable-finally. *k*, *g*, and *q* have fricative allophones before another, nonvelar stop; *g* is also fricative intervocally. *ɣ* functions as the back equivalent of *g*, and has a plosive allophone after a homorganic nasal, as in *qaŋ-ya*, dative of *qan* ‘blood’. Because of the loss of the *i/ī* opposition, the distinctions *k/q* and *g/ɣ* are phonemic even in indigenous words, e.g. *kir* ‘enter’, *qir* ‘scrape’. There is no corresponding phonemic distinction between back and front *ŋ*, the front/back value of this consonant being determined purely by its phonetic environment (front in a syllable with a front vowel, back in a syllable with a back vowel). The glottal stop is phonemic only in Arabic–Persian loans,

corresponding to Arabic *ʿayn*, e.g. *ǰāmīyāt* ‘society’ (Arabic *jamīya(t)*), although it occurs phonetically before all word-initial vowels; Arabic *hamza* after a consonant is simply lost, e.g. *māsīlā* ‘question’ (Arabic *masʿala*). The status of *ž* is somewhat marginal in indigenous words, although some dialects have initial *ž* for earlier **y* in some words, and these forms were adopted into the standard language of the former Soviet Union, e.g. *žil* (in China: *yil*) ‘year’. The phoneme *h* does not occur in indigenous words, and the opposition between *h* and *x* tends to be neutralized, in favor of *x*, immediately preceding a consonant. In loans from Arabic, Uyghur *h* corresponds to Arabic *h* and *ḥ*, while *x* corresponds to *x*, e.g. *hawā* ‘air’ (Arabic *hawā*), *hājāt* ‘need’ (Arabic *ḥāja(t)*), *xādim* ‘staff member’ (Arabic *xādim*). The pronunciation of *r* varies from trill to tap to glide, with even further realizations in nonstandard dialects.

In Chinese–Russian loanwords, the following consonants also occur: *f* (voiceless labio-dental fricative), *v* (voiced labio-dental fricative, for some speakers not distinguished from *w*), *c* (voiceless dental/alveolar affricate), *šč* (long palatal fricative or sequence of palatal fricative and affricate). It should be noted that Arabic–Persian *f* shows up as *p* in Uyghur, e.g. *ittipāq* ‘alliance’ (Arabic *ittifāq*).

45.2. Vowel harmony

Uyghur, like nearly all Turkic languages, has palatal harmony, although there are many perturbations even in indigenous words because of the merger of older front **i* and back **ī* as front *i*, and because of the similar morphophonemic behavior of *e*, the latter in part relating to the alternation of *e* with *a* or *ä*. Treating indigenous words first, the domain of palatal harmony is the word, with the proviso that the component parts of compound words are considered as separate words, e.g. *aq-teräk* ‘kind of poplar’, cf. *aq* ‘white’, *teräk* ‘poplar’.

In a given indigenous word, either all vowels will be front or all vowels will be back, except that the vowel *i*, and likewise *e* arising from umlaut (see below), may occur in both front and back vowel words. In addition, the opposition between rounded and unrounded vowels is also neutralized in non-initial syllables, usually only the unrounded vowels occurring (but see below for some complications regarding high vowels). Because of this, for the most part morphophonemically the only possible vowel opposition in non-initial syllables, including all suffix syllables, is between a high vowel

(symbolizable as *I*, e.g. in the ablative suffix *-dIn*, surfacing as *i*) and a low vowel (symbolizable as *A*, e.g. in the plural suffix *-lAr*, surfacing as *a* or *ä*). In addition, there are a few suffixes with a specifically rounded high vowel, symbolizable as *U* (e.g. in the desiderative suffix *-GU*), surfacing as *u* or *ü*. We thus find the following patterns of suffix variation.

<i>čač</i>	‘hair’	<i>čač-lar</i>	<i>čač-tin</i>	<i>qat</i>	‘harden’	<i>qat-qu</i>
<i>toy</i>	‘feast’	<i>toy-lar</i>	<i>toy-din</i>	<i>qoy</i>	‘put’	<i>qoy-yu</i>
<i>til</i>	‘tongue’	<i>til-lar</i>	<i>til-din</i>	<i>qiy</i>	‘cut’	<i>qiy-yu</i>
<i>quš</i>	‘bird’	<i>quš-lar</i>	<i>quš-tin</i>	<i>qur</i>	‘dry’	<i>qur-yu</i>
<i>bäl</i>	‘waist’	<i>bäl-lär</i>	<i>bäl-din</i>	<i>kät</i>	‘go’	<i>kät-kü</i>
<i>söz</i>	‘word’	<i>söz-lär</i>	<i>söz-din</i>	<i>köy</i>	‘singe’	<i>köy-gü</i>
<i>čiš</i>	‘tooth’	<i>čiš-lär</i>	<i>čiš-tin</i>	<i>kij</i>	‘wear’	<i>kij-gü</i>
<i>kün</i>	‘day’	<i>kün-lär</i>	<i>kün-din</i>	<i>küt</i>	‘wait’	<i>küt-kü</i>

In the above examples, the contrast between *til* and *čiš*, or between *qiy* and *kij*, illustrates the behavior of *i* as either a back or a front vowel, respectively.

Uyghur has only a very restricted form of labial harmony, whereby vowels agree in rounding with a preceding vowel. First, labial harmony does not apply to non-high vowels, which are always unrounded in non-initial syllables in indigenous words. However, a high vowel that is inserted (alternatively, that alternates with zero in other forms of the same morpheme; see § 45.3) does undergo labial harmony, so that inserted morphophonemic *I* surfaces as *i* if the preceding vowel is unrounded, as *u* if the preceding vowel is back rounded, and as *ü* if the preceding vowel is front rounded. Compare the following examples with the first person possessive suffix *-(I)m*.

<i>kiši</i>	‘person’	<i>kiši-m</i>
<i>čač</i>	‘hair’	<i>čeč-im</i>
<i>toy</i>	‘feast’	<i>toy-um</i>
<i>til</i>	‘tongue’	<i>til-im</i>
<i>quš</i>	‘bird’	<i>quš-um</i>
<i>bäl</i>	‘waist’	<i>bel-im</i>
<i>söz</i>	‘word’	<i>söz-üm</i>
<i>čiš</i>	‘tooth’	<i>čiš-im</i>
<i>kün</i>	‘day’	<i>kün-üm</i>

(For the vowels in the items ‘hair’, ‘waist’, see the discussion of umlaut below.) Compare also the inserted vowel between the final two consonants of the following stems, presented together with third person possessive suffixed forms (with the suffix *-(s)I*, where the vowel is not inserted, and where there is therefore no labial harmony).

<i>ogul</i>	‘son’	<i>ogl-i</i>
<i>köñül</i>	‘heart’	<i>köñl-i</i>

There is some variation, even among standard speakers, concerning the exact extent of labial harmony. Thus, some speakers, and the written language, do not apply labial harmony to an inserted vowel if the morpheme in question already contains an unrounded vowel, e.g. first person plural possessive *quş-imiz* ‘our bird’, whereas other speakers have *quş-umiz*. Some speakers apply labial harmony to high vowels more generally, either to all high vowels not resulting from raising (thus, *quş-umuz*), or to all high vowels whatever their origin, and some of this variation is tolerated within the spoken standard. In some cases, the fact that a high vowel in a suffix is treated as inserted or not is arbitrary from a synchronic viewpoint; thus, one simply has to learn that the adjectival suffix *-IIG* undergoes labial harmony, as in *süt-lük* ‘milky’, from *süt* ‘milk’.

Before the merger of earlier **i* and **ĩ*, in indigenous words *k* was in complementary distribution with *q*, and *g* was in complementary distribution with *ɣ*: *k* and *g* occurred in front-vowel words, *q* and *ɣ* in back-vowel words. With the merger of the two high unrounded vowels, however, a taxonomic phonemic opposition was created between front and back velars; indeed, there are minimal pairs, e.g. *kiy* ‘wear’, *qiy* ‘cut’. Thus, a taxonomic phonemic representation of Uyghur, even if restricted to indigenous words, must distinguish between *k* and *q*, and between *g* and *ɣ*. In words of Arabic–Persian origin, however, back *q* and *ɣ* occur freely with front vowels, representing Arabic *qāf* and *ḡayn* respectively, as in *qālām* ‘pen’ (Arabic *qalam*), *yārbīy* ‘western’ (Arabic *ḡarbīy*). It is also possible for front *k* and *g* to occur in back-vowel environments, e.g. *gal* ‘throat’ (cf. Persian *galū*). Given the taxonomic phonemic status of the *k/q* and *g/ɣ* oppositions, there are alternations between front and back velars in suffixes corresponding to those between front and back vowels, with morphophonemes *K* (*k/q*) and *G* (*g/ɣ*). Thus the dative suffix *-GA* has a front allomorph in *biz-gä*, dative of *biz* ‘we’, but a back allomorph in *u-lar-ya*, dative of *u-lar* ‘they’, itself plural of *u* ‘he, she’; likewise the denominal verbal suffix *-(I)K* has variants *-ik* and *-iq* (as well as *-uq* and *-ük*—see above—and *-k* after a vowel), e.g. *bir-ik* ‘unite’ (cf. *bir* ‘one’), *yol-uq* ‘meet’ (cf. *yol* ‘way’). However, if such a suffix occurs after an ‘exceptional’ velar, i.e. a front velar in a back-vowel word or a back velar in a front-vowel word, then the quality of that preceding velar determines the quality of the initial velar of the suffix, e.g. *taktak* ‘clapper’ (cf. Persian *tak* ‘blow’), dative *taktak-ka*, *xāliq* ‘people’ (Arabic *xalq*), dative *xāliq-qä*;

note that in these examples, the front/back quality of the velar in the suffix does not agree with the front/back quality of the vowel.

As in other Turkic languages, vowel harmony in Uyghur operates from left to right. In indigenous words, this can be seen most clearly in compounds, where it is the last component that determines the vowels of suffixes, e.g. *aq-teräk* ‘kind of poplar’, dative *aq-teräk-kä*. It is also seen in loanwords, both Arabic–Persian and Chinese–Russian, which violate palatal harmony stem-internally, and whose last vowel determines the quality of suffixal vowels, e.g. *hājät* ‘need’ (Arabic *hāja(t)*), plural *hājät-lär*; *qähwa* ‘coffee’ (Arabic *qahwa*), locative *qähwida*, *däšö* ‘university’ (Chinese *dàxué*), plural *däšö-lär*.

Clitics (see § 45.4) in Uyghur divide into two subclasses: those that undergo vowel harmony and those that do not. Many clitics which undergo vowel harmony in other Turkic languages do not do so in Uyghur. Thus, Uyghur interrogative *-mu* does not undergo vowel harmony, in contrast to Turkish *mI*, Tatar *-mE*, as in Uyghur *käl-di-m-mu*, Turkish *gel-di-m mi* ‘did I arrive?’ (‘come-past-first:person:singular-interrogative’).

Two other phonological processes affecting vowels are also conveniently dealt with here, namely umlaut and raising, both of which are characteristic of Uyghur, in particular its Northern dialects; it should be noted that raising in Uyghur is a much more widespread phenomenon than the highly restricted like-named phenomena of Turkish and Tatar.

By umlaut, the short vowel *a* or *ä* of an initial open syllable is raised (and, in the case of *a*, fronted) to *e* if the next following vowel is *i*, whether corresponding to earlier **i* or **ĩ*, whether or not inserted—but not *i* resulting from raising (see below). Thus, adding the third person possessive suffix *-(s)I* and the first person singular possessive suffix *-(I)m* gives forms as follows:

<i>baş</i>	‘head’	<i>beš-i</i>	<i>beš-im</i>
<i>tär</i>	‘sweat’	<i>ter-i</i>	<i>ter-im</i>
<i>yāz</i> (Persian <i>yāz</i>)	‘goose’	<i>yāz-i</i>	<i>yāz-im</i>
<i>käsp</i> (Arabic <i>kasb</i>)	‘specialty’	<i>käsp-i</i>	<i>käsp-im</i>
<i>säpär</i> (Arabic <i>safar</i>)	‘journey’	<i>säpir-i</i>	<i>säpir-im</i>

In the first two items, umlaut takes place. In the third item it fails to apply, because the vowel is long. In the fourth item it fails to apply because the initial syllable is closed. In the last item, the vowel of the second syllable does not undergo umlaut because it is not the vowel of the initial syllable (although it does undergo raising—see below), while the first vowel does not undergo umlaut because the *i* of the following syllable is the result of raising.

In many lexical items, there is no synchronic alternation between *e* and *a* or *ä*, e.g. *beliq* ‘fish’ (Old Turkic *balıq*; the *i* of the second syllable is not an inserted *I*).

By raising, the short vowels *a* and *ä* in a non-initial, open syllable are raised to *i* if not in word (or phonological-phrase) final position. Thus, the plural of *töpä* ‘peak’ is *töpi-lär*, its locative is *töpi-dä*, which when followed by *tur-i-dū* ‘(s/he) lives’ (for morphophonemic *tur-A-dū* ‘live-present-third: person:singular’) gives *töpi-di tur-i-dū*. Compare, for a back-vowel word, *bala* ‘child’, plural *bali-lar*. Raising is obligatory and represented orthographically within the word; across word boundaries, occurrence of raising varies with such factors as rate of speech, and is not represented orthographically.

For purposes of vowel harmony, *i* and *e* arising from *a* and *ä* are treated as back vowels if arising from *a* and as front vowels if arising from *ä*. Thus contrast the progression *at* ‘horse’, first person singular possessive *et-im*, locative *et-im-da*, with *ät* ‘meat’, *et-im*, *et-im-dä*, where the distinction between back-vowel and front-vowel words remains intact even though there is a phonetic merger between the middle items of each triad. In Chinese–Russian loans, *e* can function as a back vowel, e.g. *Xebey-da*, locative of *Xebey* ‘Hebei’ (Chinese *Héběi*).

45.3. Syllable structure and boundary phenomena

The basic syllable structure of Uyghur is (C)V(C). Syllables with an initial vowel occur only word-initially, and since there is phonetically a glottal stop preceding such a word-initial vowel, one might argue that an initial consonant is obligatory, thus reducing the range of possibilities to CV(C). Sequences of vowels are thus prohibited, and a glottal stop is inserted to break up such hiatuses in loanwords, e.g. *ukrāʔna* ‘Ukraine’. Word-initially, *ŋ* is excluded. Syllable-finally, the *ʔ* is excluded; thus, Arabic *ḏū l-qaʔda*, the eleventh month of the Islamic calendar, appears in Uyghur as *zulqädä*. The status of final consonant clusters is debatable. In most instances where one might expect a final consonant cluster, for instance because of a morpheme-final consonant cluster occurring before a vowel-initial suffix (e.g. *oyl-i* ‘his son’, i.e. ‘son-third:person’, *ism-i* ‘his name’), the vowel *I* is inserted when the cluster occurs syllable-finally, as in *oyul*, plural *oyul-lar*, *isim* (Arabic *ism*), plural *isim-lär*. If the first consonant is a liquid or *y*, a further possibility is for this first consonant to be dropped and the preceding vowel lengthened in com-

pensation, as with *tōt* ‘four’, alongside *tört* (the latter form being used in the former Soviet Union). Sometimes both possibilities are found with the same word, e.g. *xāliq* or *xāq* ‘people’ (Arabic *xalq*) (cf. *xālq-i* ‘his people’). See below for the special treatment of syllable-final *ŋg*. However, one does probably have to reckon with the existence of some final CC clusters even in native words, the first consonant being a liquid or *y*, the second a stop, as in *türk* ‘Turk’, *išlā-yt-ti-m* ‘I used to work’, i.e. ‘work-habitual-past-first:person:singular’, and a few more in words of Arabic–Persian origin, e.g. *kāsp* ‘specialty’ (Arabic *kasb*). In Russian loanwords initial and final consonant clusters are often implied by the orthography, but in practice initial clusters are broken up by vowel insertion in pronunciation, and final clusters are treated likewise, e.g. *kulup* ‘club’ (Russian *klub*), *istātīstika* ‘statistics’ (Russian *statistika*), *Mīniski* ‘Minsk’. In Chinese, final consonant clusters do not occur, and initial clusters are highly restricted (to consonant plus *w* or *y*), although they do show up in Uyghur, e.g. *dyānshī* ‘television’ (Chinese *diànshì*).

In addition to these constraints on syllable structure, there are two other important phonotactic constraints. First, word-final stops can only be voiceless (and *ɣ* is likewise excluded in favor of *q*). Second, clusters of obstruents must be either all voiced or all voiceless.

Various phonological alternations have, at least in part, their basis in the phonotactic constraints just discussed.

A number of stems end in a voiceless stop word-finally, and more generally syllable-finally, but end in the corresponding voiced stop when preceding a vowel, as in the following examples with and without the dative suffix *-GA* and the third person possessive suffix *-(s)i*: *mäktäp* ‘school’ (Arabic *maktab*), *mäktäp-kä*, *mäktib-i*, *zawūt* ‘factory’ (Russian *zavod*), *zawūt-qa*, *zawūd-i*, *išäk* ‘donkey’, *išäk-kä*, *išig-i*, *ayaq* ‘foot’, *ayaq-qa*, *ayiy-i*. Other words, however, show a final voiceless stop throughout: *tōp* ‘ball’, *tōp-ka*, *tōp-i*, *at* ‘name’, *at-qa*, *et-i*, *aq* ‘white (one)’, *aq-qa*, *eq-i*. Note that in principle words of both indigenous, Arabic–Persian, and Russian origin can occur in either set (Chinese does not provide a stop-final model), although there are clear statistical skewings, most noticeably that the alternation of voiceless and voiced in indigenous words is frequent only with velars. Overall, however, given the existence of these two sets, and the absence of a third set that would have voiced stops throughout (following from the ban on word-final voiced stops), a straightforward analysis is to assume that the stems of the first class end morphophonemically in a voiced stop, which is devoiced syllable-finally. Words of the second class end morphophonemically in a voiceless stop, which thus undergoes no alternation.

In some words, word-final *g* turns into the corresponding fricative, rather than being devoiced, e.g. *bäg* ‘bey’ is pronounced with a final front velar fricative allophone. There is one further slight twist to the picture concerning word-final devoicing. Some stems end in *-ŋg*, e.g. *räng-* ‘color’ (cf. *räng-i* ‘his color’). Syllable-finally, the *g* simply drops, rather than devoicing, to give *räng* ‘color’, *räng-lik* ‘colorful’.

Constraints on final consonant clusters give rise to alternations between stem-final CIC in syllable-final position and CC prevocally, as already illustrated with the example *oyul* ‘son’, plural *oyul-laŋ*, but *oyl-i* ‘his son’. This includes words of Arabic–Persian origin, e.g. *xäliq* ‘people’ (Arabic *xalq*), *xälq-i*. Since some words have a stable *I* in this position, e.g. *beliq* ‘fish’, *beliy-i* ‘his fish’, the most elegant analysis is to assume that in words like *beliq* the *I* is present morphophonemically, while in words like *oyul* it is absent morphophonemically, being inserted when the consonant cluster (in this case, *yl*) would otherwise be syllable-final. An analysis with vowel deletion rather than vowel insertion would fail to distinguish the two classes in a principled way. With some loanwords ending in a consonant cluster prevocally, the cluster is simplified in a more idiosyncratic way syllable-finally; thus *dost-* ‘friend’ ((Early Modern) Persian *dōst*), is pronounced in isolation as *dos*, but contrast *dost-um* ‘my friend’.

Constraints against changes in voicing within obstruent clusters give rise to the following phenomena. Many suffixes and clitics begin basically with a voiced stop, as can be seen in their behavior after non-obstruents, such as vowels. After a voiceless obstruent, these same suffixes begin with the corresponding voiceless obstruent. Compare the behavior of the suffixes *-dA* (locative) and *-GA* (dative) in the following items:

<i>bala</i>	‘child’	<i>bali-da</i>	<i>bali-ya</i>
<i>at</i>	‘horse’	<i>at-ta</i>	<i>at-qa</i>
<i>köz</i>	‘eye’	<i>köz-dä</i>	<i>köz-gä</i>
<i>čüş</i>	‘dream’	<i>čüş-tä</i>	<i>čüş-kä</i>

The opposite direction of voice assimilation is found where a voiced fricative occurs before a suffix or clitic beginning with a voiceless obstruent, as here the fricative is devoiced, e.g. *söz* ‘word’, *sös-siz* ‘unquestionably’, *taŋ* ‘mountain’, *tax-či* ‘itinerant trader in mountainous area’; this assimilation is not shown in the orthography. *z* before a voiceless obstruent other than *s* is only partially devoiced, as in *ingilíz-čä* ‘English-manner, in English’.

A number of more specific assimilations exist, which find only sporadic reflection in the orthography (usually only morpheme-internally). Stops

(including affricates) and fricatives in the dental to palatal regions assimilate completely to a following *s*; with *sös-siz*, discussed in the previous paragraph, one can compare *küč* ‘power’, *küs-siz* ‘powerless’. *n* assimilates to a following *l*, e.g. *Turpan* ‘Turfan’, *turpal-liq* ‘Turfan person’. It also assimilates in place of articulation to a following stop, e.g. *on* ‘ten’, *bir* ‘one’, but *om bir* ‘eleven’, *qan* ‘blood’, dative *qan-ya*.

A number of suffixes have alternative forms, one beginning with a consonant and used after a vowel, the other beginning with a vowel and used after a consonant. For those suffixes that have a variable initial vowel, this vowel is nearly always *I*, which, as an inserted vowel, undergoes (or at least can undergo) labial harmony. Compare the possessive suffixes of the first person singular (*-(I)m*) and of the first person plural (*-(I)miz*) below.

<i>kiši</i>	‘person’	<i>kiši-m</i>	<i>kiši-miz</i>	<i>kiši-si</i>
<i>čiš</i>	‘tooth’	<i>čiš-im</i>	<i>čiš-imiz</i>	<i>čiš-i</i>

The suppositional future (aorist) suffix is unusual in having a low variable vowel, i.e. *-(A)r*, as in *oqu* ‘read’, *oqu-r*, *bar* ‘go’, *bar-ar*, *kör* ‘see’, *kör-er*. For those suffixes with a variable initial consonant, this is usually *y*, but a few suffixes have some other consonant, e.g. third person possessive *-(s)I*, as illustrated above. Variable *y* can be illustrated with the potential suffix *-(y)Al*, which gives, when added to *oqu* ‘read’, *oqu-yal*, when added to *išlä* ‘work’, *išli-yäl*, but *bar* ‘go’, *bar-al*. It should be noted that while in some instances these variable segments clearly avoid non-phonotactic sequences, such as **čiš-m*, **kiši-i*, in other cases their use is more generalized: there is thus nothing non-phonotactic about **čiš-miz*.

After a long vowel, both variable *y* and variable *I* are inserted, e.g. *bahā* ‘price’ (Persian *bahā*), *bahā-yim* ‘my price’, *dāšō* ‘university’ (Chinese *dàxué*), *dāšō-yüm* ‘my university’, although the first person singular possessive suffix is usually *-(I)m*. No doubt because of the frequent phonetic identity of items like *bahā* ‘price’ and *bahār* (Persian *bahār*) ‘spring’, both [bahā], one also finds a parallel to British English ‘intrusive *r*’, with *bahār-im* as an alternative for ‘my price’, in addition to meaning ‘my spring’; however, this morphological merger is not complete, since with the third person possessive suffix ‘his price’ can only be *bahā-si*, while *bahār-i* is unequivocally ‘his spring’.

45.4. Stress

As a general rule, stress in Uyghur falls on the final syllable, and the addition of suffixes will cause the stress to move further to the right, e.g. *kiší* ‘person’,

plural *kīši-lár*; first person possessive *kīši-lir-ím*. Stress is realized as a pitch higher than that of adjacent syllables, although preceding a pause this is realized as a fall from high to low pitch. However, there are various departures from this basic pattern, some of them of different nature from the exceptions found in Turkish or Tatar.

First, it is necessary to make a distinction between light and heavy syllables. Light syllables have the structure (C)V̆, i.e. are open syllables containing a short vowel. Heavy syllables have the structure, (C)V̆, (C)VC, or (C)VCC, i.e. either have a long vowel or are closed (or both). If the final syllable of a word is light and the penultimate syllable is heavy, then stress falls on the penultimate, e.g. *yáxşı* 'new', *māšína* 'machine' (Russian *mašina*), but the locative of the last word is *māšini-dá*, because none of the heavy syllables is the penultimate syllable. It follows from this formulation that if the last two syllables are both heavy, the stress will again fall on the last syllable, e.g. *ōzám* 'myself' ('self-first:person:singular'). If a strong syllable does not receive primary stress, then it may receive secondary stress, i.e. a more accurate representation of some preceding examples is *ōzám*, *māšína*, *māšini-dá*, or indeed *òzám*, *màšína*, *mašini-dá*, since, as noted in § 45.1, vowel length is not normally realized phonetically in unstressed syllables. As the last example illustrates, only the last heavy syllable in a string of heavy syllables lacking primary stress will bear secondary stress. It follows from this that in words that have non-final stress in their citation form, the stress may nonetheless move to the right when a suffix is added, unlike the situation with lexical non-final stress in Turkish. Many words cited in other sources as exceptions to final stress are accounted for by Hahn (1991) in terms of vowel length.

As in other Turkic languages, it is necessary to distinguish between suffixes (which may attract primary stress) and clitics (which fail to attract primary stress and do not allow primary stress to pass to their right). Some clitics are partially integrated phonologically into the preceding word other than in terms of stress, for instance being subject to vowel harmony; Hahn (1991: 19, 28) calls these 'prestressed suffixes'. One such prestressed suffix is negative *-mA*, as in *kál-mi-di* 's/he did not come' ('come-negative-past'). (Exceptionally, in the present-future, *-mA* is a suffix, e.g. *bol-má-y-män* 'I am not' ('be-negative-present:future-first:person:singular').) Others do not undergo vowel harmony: interrogative *-mu* and the predicative suffixes, such as first person singular *-män* and second person singular polite *-siz*, are non-harmonic clitics, as can be seen in *yáxşı-mu-siz* 'are you well? hello!'

where the domain for primary stress assignment is *yaxši-*, for secondary stress assignment the two domains *-mu* (which receives no stress as it is a weak syllable) and *-siz* (which is a strong syllable and therefore receives secondary stress). In *uyğur-či-din* ‘from the Uyghur language’ (‘Uyghur-manner-ablative’), *-čA* is a harmonic clitic while *-dIn* is a suffix, and the word is divided into two parts for purposes of stress assignment, primary stress being assigned to *uyğur-*, secondary stress to *-či-din*. The boundary between suffixes and clitics is subject to some idiolectal variation: thus, agentive *-čI*, adjectival *-IIG*, and privative *-sIz* are treated by some speakers as suffixes and by others as clitics, thus giving rise to stress variants like *tay-líq/ táy-liq* ‘mountainous’, from *tay* ‘mountain’. Likewise, there are cases intermediate between harmonic and non-harmonic suffixes: for instance, the suffix glossed above as ‘manner’ and which also serves to form diminutives is basically a non-harmonic clitic *-čä*, e.g. *uyğur-čä* ‘Uyghur language’, *kitāp-čä* ‘booklet’, *kitāp-čä-m-dä* ‘in my booklet’, but a harmonic clitic *-čA* when raised to *-či*, as can be seen from the fact that the front/back quality of a following vowel is determined by that of the vowel preceding *-čA*, as in *kitāp-či-da*, locative of *kitāp-čä*.

The only explicit statement I have seen on compound stress in Uyghur (Hahn 1991: 28–29), illustrated with a coordinative compound, is that, in a pattern the inverse of that of, for instance, Turkish, primary stress falls on the second component, e.g. *änčä-múnčä* ‘somewhat, occasionally’ (‘like:that-like:this’; the two occurrences of *-čA* are the clitic referred to in the preceding paragraph). *Izafet* (possessive) compounds seem to have primary stress on the first component, as in Turkish, e.g. *tik-ış maşini-si* ‘sewing machine’ (‘sew-nominal machine-third:person’). Stress in Uyghur clearly requires further investigation.

References

- Grimes, Barbara F., ed. 1992. *Ethnologue: Languages of the World*, 12th ed. Dallas, Tex.: Summer Institute of Linguistics.
 Hahn, Reinhart F. 1991. *Spoken Uyghur*. Seattle: University of Washington Press.

This page intentionally left blank.

7 Caucasian Languages

This page intentionally left blank.

Georgian Phonology

Howard I. Aronson

University of Chicago

Georgian is the only written member of the Kartvelian (South Caucasian) linguistic family, having its own alphabet. It and the remaining languages of this family, Svan, Mingrelian, and Laz, share a number of phonological features in common with the other autochthonous languages of the Caucasus (Northwest Caucasian or Abkhaz-Adyghe and Northeast Caucasian or Nakh-Daghestanian), though no genetic relationship has been established between these groupings. Most striking among these is a three-way opposition in stops, contrasting voiceless aspirate, voiceless glottalized, and voiced. (This three-way opposition is also found in Eastern Armenian and Ossetian.)

46.1. Consonants

Literary Georgian has 28 consonantal phonemes. As typical for the autochthonous languages of the Caucasus, Georgian distinguishes three series of stops: (a) voiceless aspirated, (b) voiceless glottalized, and (c) voiced, although the last series is in fact rather weakly voiced, leading one to perhaps assume a basic opposition of tense/lax rather than voiceless/voiced. The stop series are as in Table 46-1.

Table 46-1. Georgian Stops

Labial			Dental/Alveolar			Alveopalatal			Velar		
a	b	c	a	b	c	a	b	c	a	b	c
p	p̥	b	t	t̚	d				k	k̚	g
			c	ç	ʒ	č	č̚	ž			

The stops /t, t̚, d/ are dental, while the affricates /c, ç, ʒ/ are alveolar. Fricatives are basically restricted to the Dental and Alveopalatal series and show a two-way opposition of (a) voiceless and (b) voiced, as in Table 46-2.

Table 46-2. Georgian Fricatives

Alveolar		Alveopalatal	
a	b	a	b
s	z	š	ž

With the exception of the postvelar series (see § 46.1.1), voicing is not distinctive for the remaining consonant phonemes. The liquids, nasals, and glides are as in Table 46-3.

Table 46-3. Georgian Liquids, Nasals, and Glides

Labial	Dental	Glottal
m	n	
	r	
	l	
v		h

/v/ functions very much like a glide. It is generally pronounced with very little friction and shows a wide variety of phonetically conditioned variants. Initially before vowels and after voiced consonants it is generally realized as [v, w, β], before voiceless consonants as [f, φ], and after voiced consonants as [hv] (voiceless [w]) or [φ]. It can cause marked rounding of a preceding consonant; see below.

/l, r, m, n/ are generally voiceless initially before voiceless consonants and between voiceless consonants: *mšia* ‘I’m hungry’ [mʰšia], *lxini* [lʰxini] ‘feast’, *rka* [rkʰá] ‘horn’, *nčkrevá* [nčkʰréva] ‘quiver (of liquid)’.

/r/ is generally a single flap. When surrounded by voiceless consonants it tends to become devoiced or even lost: *prtxili* ‘cautious’ [pʰrtxíli] or [ptxíli]. Between voiced consonants, also, there is a tendency for its loss: *brzandebit* ‘you are’ [brzandebitʰ] or [bzandebitʰ].

/l/ is generally somewhat velarized ([ɫ]), except before the vowel /i/.

/h/ in literary Georgian is limited to one grammatical morpheme (in general, the marker of a third person dative before the consonants *p, k, k, g, q*), to relatively recent loanwords (e.g. *higiéna* ‘hygiene’, *himni* ‘hymn’), and to interjections (e.g., *ha-žan!* ‘bravo’, *heri!* ‘forward!’). In the very few native words in which it occurs, it is often replaced by *χ* (e.g., *ho ~ xo* ‘yes’, *hoda ~ xoda* ‘as I was saying’) or is completely lost (e.g., *haķaraķi ~ aķaraķi* ‘stupid’).

46.1.1. *Postvelar*

The postvelar series in contemporary literary Georgian consists of fricatives, but these behave functionally the same as stops, showing the same three-way opposition of aspirated, glottalized, and voiced, as in Table 46-4. (For simplicity's sake, these will be represented by the symbols *x*, *q̣*, and *ǵ* respectively.)

Table 46-4. Georgian Postvelars

Postvelar		
a	b	c
χ	χ̣	γ

The pronunciation of the sound represented by *χ̣* (*q̣*) varies from a glottalized postvelar stop [qʔ] through an affricate [q̣χʔ] to a simple glottalized fricative [χ̣].

46.2. *Vocalism*

Georgian has a 5-vowel system, opposing three degrees of height, and front-unrounded vs. back-rounded, as in Table 46-5. To the left of the phonetic symbols are given the corresponding vowel-letters of the Georgian alphabet.

Table 46-5. Georgian Vowels

i	u	o	ɔ̃
ε	ɔ	ɣ	ɯ
ɑ		ɔ̄	

Uturgaize (1976: 11) gives the acoustic distinctive features for the vocalic system shown in Table 46-6.

46.3. *Stress*

Stress is clearly non-distinctive, although there is a good deal of controversy as to where stress falls in the word. Some view stress as penultimate and others as antepenultimate. No other prosodic features are distinctive in

Table 46-6. Georgian Vocalic Distinctive Features

	α	ε	i	$\mathfrak{O}$	u
Compact	+	+	−	+	−
Flat	−	−	−	+	+
Grave	+	−	−	+	+

Georgian. Georgian is basically a syllable-timed language, with minimal variation in vowel strength depending on location in the word.

Modern literary Georgian has no diphthongs; e.g., words such as *caiḡeb* ‘you will take’, *žoḡei* ‘jockey’, *tramvai* ‘streetcar’ are trisyllabic and *vaia-raḡeb* ‘I arm someone’ has five syllables. Nasals and sonants are never syllabic, e.g., *grznobs* ‘feels’, *mgavs* ‘resembles me’, *lxins* ‘joy (dative)’, *rka* ‘horn’, *gvprcxvnis* ‘is peeling us’ are all monosyllabic, and forms such as *ndoma* ‘will, desire’, *mdivans* ‘secretary (dat.)’ are bisyllabic.

46.4. The phonemic inventory

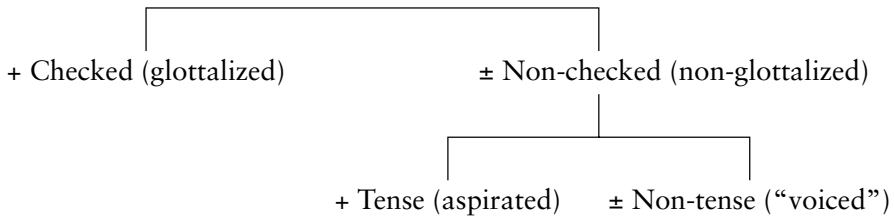
Table 46-7. Georgian Phonemic System

[illegible]

Modern Georgian has the phonemic system shown in Table 46-7. Below the phoneme symbol is given the corresponding letter of the Georgian alphabet.¹ (For the vocalic system, see Table 46-5.)

46.5. Hierarchy

In Georgian the glottalized stops (including /q/) are marked with respect to the corresponding aspirates. Within the glottalized, /p/ is the most marked; its occurrence in native Kartvelian roots is extremely rare. One possible interpretation of the system of stops has the following hierarchy:



46.6. Dialects

Georgian dialects differ little from the literary language in phonemic inventory. Mountain dialects tend to have an additional postvelar voiceless aspirated stop q^h , which merged with χ in the literary language. Dialects also show the glide y [j], which forms diphthongs; in the literary language y has been lost in all environments. These two additional phonemes, q^h and y , were found in the Old Georgian literary language.

46.7. Automatic (phonologically conditioned) alternations

Voiced stops in word-final position become aspirated, e.g., *keteba* ‘doing’ but *a^hketep* ‘you (sing.) do’. Normally /v/ disappears before the rounded vowels /u, o/, e.g., *gamoacxv-e* ‘you baked’, but *gamoacx-o* ‘he baked’, [va^hket^hep^h/a^hket^hep^h] ‘I/you make it’ but [u^hket^hep^h] (< *v-u^hketeb*) / [u^hket^hep^h] ‘I/you make it for them’. /v/ often disappears also before and after labials,

1. The order of the letters of the Georgian alphabet is:

ა ბ გ დ ე ვ ზ თ ი კ ლ მ ნ ო პ ზ რ ს ტ უ ფ ქ ც ღ ჩ ც ც ჯ ხ
a b g d e v z t i k l m n o p ო რ ს ტ უ ფ ქ ც ღ ჩ ც ც ჯ ხ

though there are exceptions to this: *svams* ‘drinks’, but *sma* (< *sv-m-a*) ‘drinking’; *ambavi* ‘thing (nom.)’, but *ambis* (gen.); but compare *krtamavs* ‘bribes’ with the verbal noun *krtamva* ‘bribing’. Metathesis of liquids and nasals with *v* in sequences of the type CN*v*-, CR*v*- is also common: *xn-av-s* ‘plows’, but *xvn-a* ‘plowing’; *kl-av-s* ‘kills’, but *ḵvl-a* ‘killing’; *ḵr-av-s* ‘pastes’, but *ḵvr-a* ‘pasting’.

The normative rules of pronunciation do not allow assimilation between consonants, although such assimilation is clearly heard in the speech of educated Georgians and is even attested in variant spellings; cf. normative *šecdoma* ‘mistake’ and *Ġvtismšobeli* ‘the Virgin Mary’ (from *Ġmert* ‘God’) alongside the variant forms *šectoma* and *Xvtismšobeli*.

46.8. Major grammatically conditioned alternations

The most common alternation in the nominal system is **syncope**, the loss of stem vowels /e, a, o/ in the genitive, instrumental, and adverbial cases and throughout the plural in *-eb-*. Syncope is most common when the consonant following the vowel is a nasal, liquid, or /v/. The more complex the resulting cluster, the more likely syncope is to occur; cf. *ḵari* ‘door’, genitive *ḵaris*, but *ciskari* ‘dawn’ (lit. ‘heaven’s-door’), gen. *ciskris*; *kali* ‘woman’, gen. *kalis*, but *mġvdeli* ‘priest’, gen. *mġvdli*s. When the alternating vowel is /o/, it is replaced by /v/ if not in the vicinity of a rounded vowel or a labial consonant. Examples:

	‘window’	‘windows’	‘walnut’	‘walnuts’
Nom	<i>panžara-</i>	<i>panžr-eb-i</i>	<i>nigoz-i</i>	<i>nigvz-eb-i</i>
Erg	<i>panžara-m</i>	<i>panžr-eb-ma</i>	<i>nigoz-ma</i>	<i>nigvz-eb-ma</i>
Dat	<i>panžara-s</i>	<i>panžr-eb-s</i>	<i>nigoz-s</i>	<i>nigvz-eb-s</i>
Gen	<i>panžr-is</i>	<i>panžr-eb-is</i>	<i>nigvz-is</i>	<i>nigvz-eb-is</i>
Instr	<i>panžr-it</i>	<i>panžr-eb-it</i>	<i>nigvz-it</i>	<i>nigvz-eb-it</i>
Adv	<i>panžr-ad</i>	<i>panžr-eb-ad</i>	<i>nigvz-ad</i>	<i>nigvz-eb-ad</i>

(Cf., however, *potoli* ‘leaf’, *oboli* ‘orphan’, which have the genitives *potl-is*, *obl-is*, without /v/ because of the preceding /o/ and /b/.)

The stem-final vowels /a, e/ generally undergo **truncation** before the endings of the dative, genitive, and adverbial and, for /a/, throughout the plural. This is illustrated for /a/ in *panžara* above.

In the verbal system, ablaut is confined to a small group of irregular verbs. There are three patterns of ablaut:

(1) the vowel /e/ or /a/ is inserted in the 1st and 2nd person of the aorist; e.g.:

‘catch’

	Future		Aorist	
1.	<i>mo-v-čr-i</i>	<i>mo-v-čr-i-t</i>	<i>mo-v-čer-i</i>	<i>mo-v-čer-i-t</i>
2.	<i>mo- -čr-i</i>	<i>mo- -čr-i-t</i>	<i>mo- -čer-i</i>	<i>mo-v-čer-i-t</i>
3.	<i>mo- -čr-i-s</i>	<i>mo- -čr-i-an</i>	<i>mo- -čr-a</i>	<i>mo- -čr-a</i>

Similarly future 1sing. *mo-v-ḱl-av*, 3sing. *mo-ḱl-av-s*, but aorist 1sing. *mo-v-ḱal-i*, 3sing. *mo-ḱl-a* ‘kill’.

(2) the vowel /a/ is inserted in all forms of the aorist; e.g.:

‘count’

	Future		Aorist	
1.	<i>ča-v-tvl-i</i>	<i>ča-v-tvl-i-t</i>	<i>ča-v-tval-e</i>	<i>ča-v-tval-e-t</i>
2.	<i>ča- -tvI-i</i>	<i>ča- -tvI-i-t</i>	<i>ča- -tval-e</i>	<i>ča- -tval-e-t</i>
3.	<i>ča- -tvI-i-s</i>	<i>ča- -tvI-i-an</i>	<i>ča- -tval-a</i>	<i>ča- -tval-es</i>

(3) In some, though not all, verbs that take no formant in the present or future, the vowel /e/ in the future of the “transitive” becomes /i/ in the aorist and in all present, future, and aorist stem forms of the derived “intransitive”; e.g.:

	‘spread’	‘be spread’	‘spread’	‘be spread’
	Future singular		Aorist singular	
1.	<i>ga-v-pen</i>		<i>ga-v-pin-e</i>	
2.	<i>ga- -pen</i>		<i>ga- -pin-e</i>	
3.	<i>ga- -pen-s</i>	<i>ga-i-pin-eb-a</i>	<i>ga- -pin-a</i>	<i>ga-i-pin-a</i>

Verbs in *-ev* lose the /v/ before vowels in the aorist, e.g., *a-rev-s* ‘he will mix up’, but aorist *a-ri-a*. Note, e.g., *da-čer-s* ‘he will write’, aorist *da-čer-a* ‘he wrote’, without alternation.

46.9. Consonant sequences

Georgian is famous for its extremely long sequences of consonants. As an example, consider the form *gvprckvnis* ‘is peeling us’, with 8 consonants (all non-syllabic, of course) preceding one vowel. To understand the structures of such accumulations of consonants it is necessary to deal with the following concepts: (a) sequences of consonantal phonemes that behave phonologically like unit phonemes (consonants with labial release, harmonic clusters); and (b) the canonical shapes of root and prefixal morphemes. These are described below.

46.9.1. Consonants with labial release

Non-labial consonants (stops, affricates, and fricatives) very commonly occur with a following /v/, which tends to labialize the preceding consonant and function as a labial release. In terms of phonological patterning, such sequences of Cv function like simple C. Examples include: *ṭvini* ‘brain’, *ṭvali* ‘eye’, *mindvris* ‘field (gen.)’, *mačvnis* ‘yogurt (gen.)’, *cvari* ‘dew’, *zveli* ‘old’, *svams* ‘drinks’, *zviadi* ‘proud’, *čvilti* ‘harpoon’, *čveni* ‘our’, *žvari* ‘cross’, *mšvenieri* ‘beautiful’, *ḵvali* ‘trace’, *gverdi* ‘side’, *qveli* ‘cheese’, *xvadi* ‘male (animal)’, *gvino* ‘wine’.

46.9.2. Harmonic clusters

Harmonic (or *decessive*) clusters (the terms were introduced by the Georgian phonetician Giorgi Axvlediani) consist of a pre-velar stop or affricate followed (a) by a corresponding velar stop, or (b) by a corresponding post-velar. By “corresponding” is meant that the whole cluster will share a single feature of glottalization, aspiration, or voicing. Such clusters are:

	a. C ₁ C ₂ (C ₂ a velar)	b. C ₁ C ₂ (C ₂ a post-velar)
Labial	<i>pḵ, pk, bg</i>	<i>pḳ, px, bğ</i>
Dental	<i>tḵ, tk, dg</i>	<i>tḳ, tx, dğ</i>
	<i>çḵ, ck, zḡ</i>	<i>çḳ, cx, zğ</i>
Palatal	<i>čḵ, čk, žḡ</i>	<i>čḳ, čx, žğ</i>

Harmonic clusters, both of type (a) (velar) and type (b) (post-velar), have a single release—glottalized, aspirated, or voiced. In terms of phonological patterning, these clusters function like a single C.

Harmonic clusters, in that they function as single consonants, can also have a labial release. Since the second member of the harmonic cluster is a velar or a post-velar, clusters in which the first member is a labial can also have a labial release. Thus, the following labially released harmonic clusters are **theoretically** possible:

	a. C ₁ C ₂ v (C ₂ a velar)	b. C ₁ C ₂ v (C ₂ a post-velar)
Labial	<i>pḵv, pkv, bgv</i>	<i>pḳv, pxv, bğv</i>
Dental	<i>tḵv, tkv, dg</i>	<i>tḳv, txv, dğv</i>
	<i>çḵv, ckv, zḡv</i>	<i>çḳv, cxv, zğv</i>
Palatal	<i>čḵv, čkv, žḡv</i>	<i>čḳv, čxv, žğv</i>

Examples. Both clusters with labial and non-labial release are given:

(a)		(b)	
<i>pḱureba</i>	‘sprinkling’	<i>pḱroba</i>	‘destruction’
<i>pḱvili</i>	‘(ground) meal’	<i>pxa</i>	‘fish bone’
<i>bgera</i>	‘sound’	<i>bḡvriali</i>	‘sparkling’
<i>ṭḱbili</i>	‘sweet’	<i>siṭqva</i>	‘word’
<i>tkvi</i>	‘you said’	<i>txa</i>	‘goat’
<i>dgas</i>	‘stands’	<i>dḡe</i>	‘day’
<i>çḱvertḱi</i>	‘string’	<i>çḱali</i>	‘water’
<i>ckera</i>	‘looking’	<i>cxeni</i>	‘horse’
<i>zgera</i>	‘beating’	<i>zḡveni</i>	‘gift’
<i>çḱviani</i>	‘intelligent’	<i>çḱvirili</i>	‘squeaking’
<i>čkara</i>	‘quickly’	<i>čxubi</i>	‘dispute’
<i>žgupi</i>	‘group’	<i>žḡanmani</i>	‘old clothes’

Thus, in addition to the 28 basic consonantal phonemes of Georgian, we can add 19 consonants with labial release, 24 harmonic clusters, and (theoretically) 24 harmonic clusters with labial release, all of which can function phonologically as unit phonemes.² This means that there is a maximum of approximately 95 sequences of the type C, Cv, CC, CCv that can function as C in Georgian. Typologically, this makes Georgian phonology more than vaguely reminiscent of the phonologies of the Northwest Caucasian (Abkhaz-Adyghei) family, cf. Ubykh with 80–82 consonantal phonemes. Interestingly, some have assumed that harmonic clusters in Georgian may have developed from unit phonemes in earliest proto-Kartvelian.

46.9.3. Non-harmonic clusters

Less common in native Kartvelian roots are non-harmonic (*accessive*) clusters, where the first member is pronounced further back in the mouth than the second member.³ Some examples include *ṭba* ‘lake’, *tbili* ‘warm’, *ḱbena* ‘biting’, *qba* ‘jaw’, *gdeba* ‘throwing,’ etc.

2. In some instances, synchronic harmonic clusters historically might represent the “zero” grade of an alternation. So, for example, compare the sequence *dg* in *dgas* ‘is standing’ with the Old Georgian aorist 1sing. *videg* ‘I stood’. In other instances, the harmonic cluster appears to have functioned as a single C in Old Georgian, cf. *vtkvi* ‘I said’.

3. Often, either in other Kartvelian languages or in Old Georgian there is evidence that a vowel has been lost between the other two consonants making up the cluster; so for example, modern Georgian *ga-v-tb-i*, but Old Georgian *gan-v-ṭep* ‘I warmed myself’.

46.10. Canonical shapes of native morphemes

The clusters of modern Georgian can be explained historically in terms of the canonical shapes of proto-Kartvelian morphemes as reconstructed by Gamqrelize and Mačavariani 1965 (Gamkrelidze and Mačavariani 1982). Basically, from the beginning of the stem on, proto-Kartvelian allowed only one vowel to appear. The canonical shape of roots could be C- or CV(R)C-, where C can represent a single consonant, a single consonant plus *v*, a harmonic cluster, or a harmonic cluster plus *v*; V represents a vowel (usually *e*); and R represents the sonants *m*, *n*, *l*, *r*. Suffixes had the canonical shape -V(C). When a suffix was added to a root, one of two patterns emerged as a result of the loss of one of the Vs: either the vowel of the root was lost (e.g., CVC-V(C) > CC-V(C)) or the vowel of the suffix was lost (e.g., CVC-VC > CVC-C). The addition of a second suffix would have the same consequences, leading to either CC-C-V(C) or CVC-C-C. As an example we can take the Old Georgian aorist form (*ga-v-*)*prckwen* 'I peeled' < **p(e)rckw* plus the suffix -*en*; but in the present, the addition of the suffix -*i* caused the *e* of the first suffix to be lost: (*v-*)*prckwn*-.⁴

46.11. Principles of borrowing

Georgian tended to borrow voiceless stops in words from Eastern languages with voiceless aspirated stops, e.g., Persian *kāmār* 'belt' > Georgian *kamari* 'id.', *pañjare* 'window' > Geo. *pañžara* 'id.', *târix* 'date' > Geo. *tariği* 'id.'. However, borrowings from Greek, associated with the introduction of Christianity, show the following correspondences: Greek *φ*, *θ*, *χ* correspond to Georgian aspirated *p*, *t*, *k*, while Greek *π*, *τ*, *κ* correspond to Georgian glottalized *p̣*, *ṭ*, *ḳ*. This tradition continues in the adaptation of words from Western languages; cf. *pañtia* '(political) party', *demokratia* 'democracy', *paktori* 'factor'. As the last example shows, voiceless stops before other consonants are generally aspirated; cf. *traktori* 'tractor', *psikiatři* 'psychiatrist'. Western [c] and [č] are borrowed with the corresponding aspirates: *Čikago* 'Chicago', *civilizacia* 'civilization'.

4. The initial *gv-* in the example given above, *gvprckvnis* 'he is peeling us', is simply the object marker of the 1st person plural, which here occurs immediately before the verbal root.

Bibliography

- Artemov, V. A., and S. M. Žgenti, eds. 1959. *Fonetičeskij sbornik, I* [Phonetic collection I]. *Trudy kafedry obščego jazykoznanija*, 3. Tbilisi: Izd-vo Tbilisskogo gosudarstvennogo universiteta imeni Stalina.
- Gamkrelidze, Thomas V., and Givi I. Mačavariani. 1982. *Sonantensystem und Ablaut in den Kartwelsprachen: Eine Typologie der Struktur des Gemeinkartwelischen*. Tübingen: Narr.
- . 1965. *Sonanṯa sistema da ablaṯi kartvelur enebši: Saerto-kartveluri sṯruṯuris ṯipologia* [The system of sonants and ablaut in the Kartvelian languages: The typology of the structure of Common Kartvelian]. Tbilisi: Mecniereba [Russian summary, pp. 381–474.]
- Nebieridze, G. 1974. *Saliteraṯuro kartuli enis generaṯorul-ponologiuri modeli da misi agebis principebi* [A generative phonology model of the Georgian literary language and the principles of its construction]. Tbilisi: Tbilisis universiṯeṯis gamomcemloba.
- Robins, R. H., and Natalie Waterson. 1952. “Notes on the Phonetics of the Georgian Word.” *Bulletin of the School of Oriental and African Studies* 14: 55–72.
- Selmer, Ernst W. 1935. *Georgische Experimentalstudien*. Oslo: Dybwad.
- Tschenkéli, Kita. 1958. *Einführung in die georgische Sprache*, I, pp. xliii–lxiv. Zürich: Amirani.
- Uturgaidze, F. G. 1976. *Kartuli enis ponemaṯuri sṯruṯura* [Phonemic structure of Georgian]. Tbilisi: Mecniereba.
- . 1978. “Fonematičeskaja sistema gruzinskogo jazyka” [The phonemic system of Georgian]. *Ežegodnik iberijsko-kavkazskogo jazykoznanija* 5: 28–98.
- Vogt, Hans. 1958. “Structure phonémique du géorgien.” *NTS* 18: 5–90.
- Žgenti, Sergi. 1956. *Kartuli enis poneṯiṯa* [Georgian phonetics]. Tbilisi: Tbilisis universiṯeṯis gamomcemloba.
- . 1965a. *Kartvelur enata poneṯiṯuri saṯitxebi* [Phonetic questions of the Kartvelian languages]. Tbilisi: Ganatleba.
- [Žgenti, S. M.]. 1965b. “O ritmičeskoj strukture frazy gruzinskogo jazyka” [On the rhythmic structure of the phrase in Georgian]. *Phonetica* 12: 147–50.

This page intentionally left blank.

Chechen Phonology

Johanna Nichols

University of California, Berkeley

47.1. Introduction

There were 956,879 Chechen at the time of the 1989 census (as reported in *Cultural Survival Quarterly*, Winter 1992, p. 25), a figure that makes the Chechen the largest ethnic group of the North Caucasus and the largest indigenous Caucasian group after the Georgians. Chechen (hereafter, Ch) is spoken in a sizable area in the eastern central part of the North Caucasus: the secessionist state of Chechnia, which occupies most of the Sunzha basin, with outliers in northwestern Dagestan and northeastern Georgia.

In pre-Soviet times, Ch was occasionally written using the Arabic script (the Chechen are almost entirely Sunni Muslims, having converted to Islam in the 17th and 18th centuries). Ch has been written systematically since 1923, when an orthography in the Latin alphabet was designed for the language; it was replaced by the present Cyrillic orthography in 1938. The orthography adds only one symbol—the letter “I”—to the standard Russian alphabet. It considerably underdifferentiates the vowel phonemes but represents the consonants economically. The additional letter “I” is used to mark glottalization of consonants (e.g. тI = *tʰ*, e.g. тIай = *tʰaj* ‘bridge’), pharyngealization of voiced consonants (e.g. дI = *dʁ*, e.g. дIа = *dʁa*: ‘there’), /ɣ/ (spelled rI, e.g. rIала = *ya:la* ‘fort’), and /h/ (xI). The Russian hard sign represents the Ch glottal stop in certain positions, e.g. пхиъ = *pxiʔ* ‘5’. The Russian soft sign follows a vowel to indicate umlaut, e.g. аьлла = *älla* ‘having said’, дуьне = *dünie* ‘world’, тоьхна = *tüöxna* ‘having struck’.

Ch has a number of dialects, which differ primarily in the presence or absence of umlaut (see § 47.3.1) and its phonemic treatment; the literary language is based on the central lowlands dialect, which has the fullest implementation of umlaut and the largest vowel inventory. The Ch self-name is (singular) *Nwoxčuo*, (plural) *Nwoxčî*; the language is *nwoxčî:ⁿ mwott* (lit. ‘language of the Chechens’). The Russian and general European term Chechen comes from the name of a lowlands village, Čečana (in Russian,

Čečan-aul, a form evidently taken from the Turkic language Kumyk that is spoken to the north of Ch).

Ch belongs to the Nakh branch of the Northeast Caucasian, or Nakh-Daghestanian, language family, an indigenous Caucasian stock with several deep branches and some 30 daughter languages. To the west of Ch is its very close sister Ingush; to the southeast in Georgia is the less closely related Batsbi (or Tsova-Tush); these three comprise the Nakh branch. Among the phonological hallmarks of Nakh-Daghestanian languages are well-developed consonant systems (with uvular and lateral affricate series, and with up to five manners of articulation including voiceless aspirate, voiced, voiceless unaspirated or tense, ejective, and tense or geminate ejective), severe restrictions on consonant clusters, well-developed vowel systems (generally with umlaut and/or ablaut in inflection and derivation), and pharyngealization as a prosodic feature of lexical roots (in addition to independent pharyngeal phonemic segments in some languages). Some Daghestanian languages are reported to have tonal oppositions (for a recent cross-linguistic survey see Kibrik & Kodzasov 1990: 311ff.); Ch lacks these, except for distinctive high tone on three particles.

Russian (hereafter R) influence on Ch dates to the 16th century when the first Cossack settlements appeared on the north Caucasian plain. The Ch were among the strongest resistors to the R conquest of the Caucasus and suffered considerable loss of population and territory. From 1944 to 1956 they (along with the neighboring and closely related Ingush and two other Caucasian groups) were exiled to Central Asia and Siberia, again with great loss of population and property. In 1991 Chechnia declared its independence from Russia, although Russia has not recognized this. Despite the ethnic tensions produced by this history of genocide and oppression, most Chechen are fluent in R and many are virtually bilingual. Those who received their education in Central Asia during the years of exile—when publication in and other official use of the Ch language were forbidden—are the most clearly bilingual and some of them appear to be R-dominant. In general, however, R fluency among the Chechen appears to be not the result of enforced Russification but the latest phase in a long-standing tendency to value fluency in an important lowlands language (traditionally, Kumyk) for inter-ethnic communication in the lowlands and an important written language (traditionally, Arabic) for writing and scholarship. In the Soviet period, R had both of these functions. One consequence has been strong lexical influence of R on Ch (much of it mandated by official policy) and

some syntactic influence, but there has been almost no phonetic or phonological influence. My observation is that even R-dominant bilinguals pronounce R words with Ch phonology and phonetics when speaking Ch, though I have been told that some Chechen speak with a R accent.

Descriptions and examples below pertain to the literary language unless otherwise indicated. Examples preceded by “#” are taken from published sources and simply transliterated (underdifferentiating vowels as the orthography does); others are phonemic transcriptions unless enclosed in square brackets. Capital letters are used to label gender classes: D gender, B gender, etc.; *D-* is the regular citation form for gender-agreeing words. The hyphen indicates any morpheme boundary, except that the equals sign (=) is used for the boundary between gender prefix and agreeing word, e.g. *d=e:xa*, *j=e:xa*, *v=e:xa*, *b=e:xa* ‘long’.

47.2. Phonemic system

47.2.1. Consonants

The consonant phonemes of Ch are as in Table 47-1. The language may be described as having pharyngealized consonants as well, though pharyngealization will be analyzed here as a morphophonemic prosody and as a phonetic segment (see § 47.5.4).

Table 47-1. Chechen Consonants

	NONCONTINUANT OBSTRUENTS				CONTINUANT			RESONANTS	
	VI	Vd	Ejec	Fortis	VI	Vd	Nasal	Liquid	Glide
Labial	p	b	(pʰ)	pp			m		v
Dental	t	d	tʰ	tt				l	
Alveolar	c	ɟ	cʰ	ss	s	z	n	r	rʰ
Palatal	č	ž	čʰ		š	ž			j
Velar	k	g	kʰ	xk	x				
Uvular	q		qʰ	qq		ɣ			
Pharyngeal		ʕ			ħ				
Glottal		ʔ			h				

The voiceless stops and affricates are aspirated, except after /s/. The degree of aspiration is similar to that of English, except that in the Kisti dialect of eastern Georgia, apparently under the phonetic influence of Georgian, aspiration is audibly heavier and is preserved after /s/. The consonants shown in the “Fortis” column are those occurring in the underlying structure of roots; almost any consonant can be doubled by focus gemination, discussed in § 47.3.7, and some geminates arise over the morpheme boundary between root-final consonant and converb suffix *-na*. The fortis series is usually geminate and never aspirated (except in the Kisti dialect, where it is aspirated and the fortis consonants could be analyzed as true geminate clusters). /xk/, phonetically [xk], is the structural and etymological equivalent of **kk*. Preaspiration is not otherwise a feature of the fortis series in any dialect I have heard (though it is systematic in Ingush, where it occurs regularly after long vowels and optionally after short vowels).

Note that /n/ is alveolar, exactly like English /n/, unlike the stops /t/, /d/, /tʰ/, which are true dentals. /x/ is slightly farther back than the velars /k/, etc., and slightly farther back than R /x/, but not as far back as the uvulars. /r/ is usually articulated with a single tap. /rʰ/ is a corresponding voiceless sound, found in final position apparently only in two roots: *vworʰ* ‘7’ and *barʰ* ‘8’. The voiceless /rʰ/ is phonetically longer than plain /r/, having either a prolonged fricated release or perhaps a second tap.

/v/ has two allophones: [v] before front vowel and [w] elsewhere.

All words begin with consonants. Those written here (and spelled in the orthography) as vowel-initial in reality begin with a glottal stop (which is phonetically prominent, of evident psychological reality, and always pronounced in continuous speech and even in singing). /ʕ/ is a pharyngealized glottal stop initially and intervocalically (i.e. in positions other than post-consonantal); it could also be analyzed as a glottal stop that precedes a pharyngealized vowel (for the analysis of pharyngealization see § 47.5.4). I use the symbol “ʕ”, which as an IPA symbol properly indicates a voiced pharyngeal continuant, in phonemic transcriptions of the voiced pharyngeal phoneme (or vowel pharyngealization, depending on one’s choice of analysis) because I take it to be the most widely recognized visual symbol for the combination of voicing and pharyngealization. Also, native speakers of Ch identify this independent phoneme with vowel pharyngealization, whose phonetic manifestation can be a voiced pharyngeal continuant (with heavy pharyngealization of the following vowel): [dʕa] ‘there’.

The contrast of voiced affricates /ʒ, ʒʰ/ and fricatives /z, zʰ/ is neutralized (with only fricatives appearing) in some or most positions. In the literary

dialect it is preserved in initial position, but even here some speakers seem to have only fricatives. Examples elicited from a speaker of the literary dialect who preserves the contrast:

/ʒ/	<i>žižig</i> ‘meat’	<i>žima</i> ‘little’
/ʒʲ/	<i>žieruo</i> ‘widow’	<i>žir</i> ‘drum’ (in a Moslem cult)

47.2.2. Vowels

The vowel phonemes of Ch are as in Table 47-2. Those printed in boldface continue etymological and underlying **i*, **u*, **e*, **o*, **a*, long and short; the others are the result of umlaut (described in § 47.3.1).

Table 47-2. Chechen Vowels

i	i:	ü	ü:	u	u:
ye	ie	qö	üö	wo	uo
e	e:			o	o:
ä	ä:			a	a:

The vowel phonemes have the phonetic values indicated by the transcription, except as follows:

/uo/	=	[wɔ:], [uə], [uo]
/qö/	=	[qö], [qe], [we]
/üö/	=	[üö], [üe]
/a/	=	[ə]

“y” and “w” are used here to transcribe initial components of short diphthongs. These components are generally not phonetically different from the consonant phonemes /j/ and /v/ respectively, but they pattern phonemically as vowels (or components thereof), e.g. in triggering the automatic appearance of glottal stop before vowel-initial words. (Historically, *y* and *w* result from diphthongization of **e* and **o*, while /j/ and /v/ are inherited consonant phonemes.) Some examples of minimal and near-minimal pairs (“J class” refers to the gender class the form agrees with; only the synchronically real morpheme boundaries are shown in phonemic forms):

Morphophonemic	Phonemic	Phonetic	Gloss
<i>j-a:xi-</i>	<i>j-e:xa-</i>	[je:x(ə)]	‘long’ (J class)
<i>j-e:x-</i>	<i>j-iex</i>	[jiex-]	‘ask’ (J class)
<i>j-ett</i>	<i>jyett</i>	[jett], [jitt]	‘cow’
<i>e:c-in-a</i>	<i>yec-na</i>	[ʔyecnə]	‘having bought’

All vowels undergo considerable allophonic distortion in the vicinity of pharyngeal consonants: *i* and *e* vowels are backed somewhat, *u* and *o* vowels fronted somewhat, high vowels diphthongized (with a mid-central onset), the diphthongs /ye/, /wo/ can be metathesized to [ej], [ow], and the transition from a preceding voiced pharyngeal can slightly lengthen a vowel. Pharyngeals (and hence pharyngealization of vowels) are of the type described as epiglottalization by Kodzasov 1986: there is constriction in the epiglottal region, with the epiglottis as active articulator (it moves back and down; there is vowel distortion due to the tongue body being tensed, flattened out, and moved forward). The acoustic effect of pharyngealization (Kingston & Nichols 1986 and work in progress) is compaction of formant transitions to the following vowel (the effect often continues well into the vowel) and a delayed onset of full vowel voicing (there is murmured or other partial voicing after a voiced consonant, and very noisy aspiration after an aspirated consonant).

/ä/ and /ä:/ are treated as phonemic here, consistent with the orthography and speakers' intuitions, but on a sufficiently abstract analysis they could be removed: /ä/ is in complementary distribution with /e/ (of which it is the allophone after a pharyngealized consonant), and /ä:/ with /e:/ (of which it is the shortened allophone, occurring in closed syllables: for shortening in closed syllables see § 47.3.3). Removing /ä/ and /ä:/ from the phoneme inventory requires that they be distinct from each other—but in fact they are minimally distinct in length for most speakers, for whom /ä:/ is phonetically short by definition since it occurs in closed syllables.

All (or probably all) vowels also occur nasalized: the genitive ending, the infinitive ending, and (at least for some speakers, and perhaps with lexical restrictions) the nominative ending of adjectives impose nasalization. Since all instances of nasalization are morphologically predictable and reflect the presence of a former **n* phoneme, nasalization is written here as a superscript (which graphically resembles the **n* segment, and which follows the orthography). Nasalization is slight, though audible even in a devoiced word-final vowel.

In addition to /ye/, /ie/, /wö/, /üö/, /wo/, and /uo/, which are the result of diphthongization of original plain vowels, Ch has the following diphthongs, in which the second element continues an original consonant:

/uj/ =	[uj], [üj]	/ow/ =	[ow], [əw]
/woj/ =	[woj]	/ew/ =	[eɥ], [oj]
/aj/ =	[əj], [ej]		

A near-minimal pair contrasting a sequence of vowel plus consonant /a:v/ with the diphthong /ow/ is the following. (A transliteration of the first word as spelled by Maciev 1961 is *a:vst*; but the phonetic length of the initial vowel shows that the syllable is not closed, and in any event the consonant sequence *vst* is impossible in Ch, so I write *a:vast*.)

Morphophonemic	Phonemic	Phonetic	Gloss
<i>a:vast</i>	/ʔa:vast/	[ʔa:vst]	‘year-old fem. goat’
<i>ow</i>	/ʔow/	[ʔəw]	‘seine’

47.3. Phonological processes

47.3.1. *Umlaut*

Certain vowels assimilate to the height, frontness, and/or rounding of a following {-i} or {-u}. (Historically, this {-i} reflects both **i* and **e*, and {-u} reflects both **u* and **o*; see the discussion of neutralization in § 47.3.2. The different vowel heights produce different ablaut grades in some dialects, but not in the literary dialect described here.)

Table 47-3 shows the operation of umlaut in the inflection of verbs. The infinitive ending contains **a* and does not produce umlaut. The present tense ending is [-u] < **o*, which rounds a preceding front vowel and raises a preceding mid back vowel. The aorist and converb endings contain **i*, which fronts a preceding back vowel and raises a preceding mid front vowel. The converb, from which the vowel is often lost to create a cluster and shorten the preceding syllable, is included only where that shortening occurs and the vowel output differs from that of the aorist. (There may be no verbs with short root /ye/ or /wo/ in Ch; the only examples with phonetic short [ye] and [wo] are before geminates or consonant clusters, i.e. in closed syllables, an environment which shortens vowels and makes it impossible to determine whether the underlying root vowels are [ye, wo] or [ie, uo].)

Umlaut in Ch, then, involves three processes: rounding, fronting, and raising. That these processes are discrete will be shown in § 47.4.3, where they will be shown to pattern differently, and independently, in various dialects.

47.3.2. *Posttonic vowel neutralization*

In noun declension, posttonic short vowels all neutralize to /a/. This does not affect verb conjugation fully: see Table 47-3, where posttonic /u/ and /i/ remain. For some speakers, the present-tense {u} does merge with /a/. The other present-tense suffix {i} (pan-dialectal **e*) does merge with /a/ for all speakers (Table 47-4).

Table 47-3. Chechen Umlaut

ROOT VOWEL	INFINITIVE *-a ⁿ	PRESENT *-u	AORIST *-ira	CONVERB *-ina	GLOSS
i	ʃilla ⁿ	ʃüllu	ʃillira		‘lie’
i:	di:ca ⁿ	dü:cu	di:cira		‘say’
u	duza ⁿ	duzu	düzira		‘fill’
u:	du:xa ⁿ	du:xu	dü:xira		‘dress’
ye					
ie	dieša ⁿ	düöšu	di:šira		‘read’
ye / _CC	dyelxa ⁿ	dqölxu	dilxira		‘cry’
wo					
wo / _CC	hwotta ⁿ	huttu	hüöttira		‘stand up’
uo	tuoxa ⁿ	tu:xu	tüxira	tqöxna	‘strike’
a	mala ⁿ	molu	melira		‘drink’
a:	a:la ⁿ	o:lu	e:lira	älla	‘say’
ow	lowza ⁿ	lowzu	lewzira		‘play’

Table 47-4. Chechen Umlaut: *i*-Verbs

ROOT VOWEL	INFINITIVE *-a ⁿ	PRESENT *-i	AORIST *-ira	CONVERB *-ina	GLOSS
a	lata ⁿ	leta	letira	letta	‘stick to’
a:	la:tta ⁿ	lätta	lättira	lättina	‘stand’

47.3.3. Shortening of long vowels in closed syllables

In a closed syllable, an underlying long vowel is shortened. Most shortened vowels remain distinct from the corresponding phonemically short vowels, distinguished by quality, breathy voice (which applies to /a:/ for some speakers and in some contexts, e.g. the dative case of ‘mother’: {na:n-na}, /na:nna/, [naṇṇə]), or a subtle contrast in length. Only the basic mid vowels /ye, ie/, /we, üö/, /wo, uo/ seem to merge entirely (in the literary dialect, for speakers I have heard; the distinctions are maintained in some dialects). The following array shows the results of shortening, contrasted with underlyingly short vowels:

Long	Shortened	Corresponding short phoneme
/i:/	[i]	[ɪ]
/u:/	[u]	[ʊ]
/ie/	[ye]	[ye] (merger of short and shortened)
/üö/	[yö], etc.	[yö], etc. (merger)
/uo/	[wo]	[wo] (merger)
/e:/	[ä], [ä]	[e]
/o:/	[ɔ]	[o]
/a:/	[a]	[ə]

A near-minimal pair is /ba:rz/ [barz] ‘grave mound’ : /barza/ [bærzə] ‘dog’s name’ (the latter an ablaut derivation from *bworz* ‘wolf’). The following array contrasts three umlaut grades in a minimal pair of verb roots: *da:xk-* ‘be located, contained’ and *daxk-* ‘put’. (Here nasalization is transcribed with a tilde.)

Infinitive:	/da:xkan/	[daxkã]	/daxkan/	[dæxkẽ]
Present:	/do:xku/	[dɔxkɯ]	/doxku/	[dɔxkɯ]
Past:	/dä:xkina/	[däxkinə]	/dexkina/	[dexkinə]

Not all consonant clusters fully close syllables. /st/ and /rt/ are two that often do not (there appears to be some inter-speaker variation, and considerable inter-dialect variation, on this point). Some near-minimal pairs showing identical vowels {uo} undergoing different degrees of phonetic shortening:

Non-shortening environments	Shortening environments
<i>muostag</i> [mwastəɣ] ‘enemy’	<i>duottag</i> [dwəttəɣ] ‘friend’
<i>kuorta</i> [kwartə] ‘head’	<i>kuoč</i> [kwəč] ‘shirt’

This description of shortening pertains to the lowlands dialect. In the Kisti dialect, to judge from the speakers I have heard, there is much less vowel shortening; length is clearly audible even in closed syllables. I have no direct information on other dialects.

47.3.4. Word-final devoicing of vowels

Final short vowels are devoiced. This applies to nasalized vowels as well. The vowel is faintly audible as a whispered vowel, and nasalization is also still faintly audible. These vowels are restored to their full voiced form in singing.

47.3.5. Loss of -i- in the anterior converb ending -ina

When the root-final consonant is a single (i.e. non-fortis) consonant other than a nasal or *ɾ*, the converb ending *-ina* loses its initial vowel and appears as *-na* (with umlaut of the root vowel triggered by the dropped *-i-*). The *-n-*

then assimilates to a preceding *-l*, *-s*, or *-t* to form a geminate *-ll-*, *-ss-*, or *-tt-* (this assimilation also happens with other consonants in allegro speech and various dialects). The loss of *-i-* does not occur in verbs with root long *-i-*, a vocalism that is at least sometimes, and possibly always, due to infixation of a number or aspect marker which may originally have been consonantal and inserted before the root-final consonant, so that here too we may have reflexes of former clusters. The last form listed below in the first group (single consonants) is an exception, synchronically; all other verbs in *-x* (there are 15 others) lose the *-i-*. The Batsbi cognate has a cluster (*doipxi* // *du:pxu* ‘clothes’) which historically explains the exception. Verbs with single consonants that fail to undergo the change are very few: there is one verb in *-m-*, one in *-n-*, four in *-r-*, and the exceptional *D-u:x-* just discussed.

Infinitive Anterior converb

Forms with no loss of *-i-*:

Single consonants:

<i>ʃa:maⁿ</i>	<i>ʃe:mina</i>	‘learn’
<i>ʃa:naⁿ</i>	<i>ʃe:nina</i>	‘pour out, spill’
<i>la:raⁿ</i>	<i>le:rina</i>	‘count’
<i>D-u:xaⁿ</i>	<i>D-ü:xina</i>	‘put on, dress’

Geminate consonants:

<i>la:ttaⁿ</i>	<i>lättina</i>	‘stand’
<i>tillaⁿ</i>	<i>tillina</i>	‘put (on top)’ (plural object)
<i>D=a:qqaⁿ</i>	<i>däqqina</i>	‘take, deprive’

Clusters:

<i>lastaⁿ</i>	<i>lestina</i>	‘shake, shudder, rock’
<i>gajtaⁿ</i>	<i>gejtina</i>	‘show’ (<i>-jt-</i> is a causative suffix)
<i>lowzaⁿ</i>	<i>lewzina</i>	‘play’
<i>D=aržaⁿ</i>	<i>D=äržina</i>	‘spread out’
<i>la:xkaⁿ</i>	<i>läxkina</i>	‘chase’ (plural object)
<i>lyeč’q’aⁿ</i>	<i>lič’q’ina</i>	‘hide’

Forms with loss of *-i-*:

<i>liepaⁿ</i>	<i>liepna</i>	‘twinkle, shine’
<i>tabaⁿ</i>	<i>tebna</i>	‘sneak up’
<i>la:caⁿ</i>	<i>läcna</i>	‘catch’
<i>tasaⁿ</i>	<i>tesna</i>	‘throw’
<i>D=iešaⁿ</i>	<i>D=iešna</i>	‘read’
<i>liegaⁿ</i>	<i>liegna</i>	‘fall, leak’
<i>tuoxaⁿ</i>	<i>tüöxna</i>	‘strike’
<i>lieʃaⁿ</i>	<i>lieʃna</i>	‘destroy’? [gloss uncertain]
<i>qieħaⁿ</i>	<i>qieħna</i>	‘drag, pull’
<i>xaʔaⁿ</i>	<i>xiʔna</i>	‘sit’

Forms with loss of *-i-* and assimilation:

<i>lata</i> ⁿ	<i>letta</i>	‘stick to’
<i>D=a:t’a</i> ⁿ	<i>dät’t’a</i>	‘split, explode’
<i>D=a:la</i> ⁿ	<i>dälla</i>	‘enter, cross threshold’

47.3.6. *Ablaut*

The term “ablaut” refers to morphologically conditioned vowel alternations of specific types. In nominal paradigms the most common ablaut pattern is an alternation of some vowel (often, but not always, a long vowel) in the nominative singular with short /a/ in the oblique forms. Some examples:

Nominative	Genitive	Dative	Gloss	Ablaut type
<i>buolx</i>	<i>belxa</i> ⁿ	<i>balxana</i>	‘work’	<i>uo</i> (= * <i>o</i> :) : <i>a</i>
<i>dwog</i>	<i>dega</i> ⁿ	<i>dagna</i>	‘heart’	<i>wo</i> (<i>uo</i> ?) : <i>a</i>
<i>šwo</i>	<i>šera</i> ⁿ	<i>šarna</i>	‘year’	<i>wo</i> : <i>a</i>
<i>dig</i>	<i>dagara</i> ⁿ	<i>dagarna</i>	‘axe’	<i>i</i> : <i>a</i>
<i>c’a</i>	<i>c’i:na</i> ⁿ	<i>c’ienna</i>	‘room’	<i>a</i> : <i>ie</i> (= * <i>e</i> :)

In addition to ablaut, the effects of umlaut can be seen in the genitive case, where the ending is underlying {-in} and the root vowel is fronted or raised.

47.3.7. *Focus gemination*

In what must once have been a productive process but now often appears to be frozen into the lexicon, the last intervocalic consonant of the stem is geminated when the word is the focus of emphasis of certain kinds (often marked by the coordinating particle *?a* in its emphatic function, and occasionally by the negative particle *ca*). This process geminates any consonant. Additional examples are the following adverbs, where it is the semantics and not a formal element in the sentence that triggers the gemination.

<i>gyerga</i>	‘nearby’	<i>juq’a</i>	‘in the middle’
<i>gyergga</i>	‘right next to’	<i>juq’q’a</i>	‘right in the middle’

47.4. Dialects. Ingush

47.4.1. *Introduction*

Chechen linguists recognize a number of major dialects, some with local varieties. Speakers I have consulted, and most published sources, report that there is good to excellent mutual intelligibility between all dialects of Ch. These dialects generally correspond to traditional tribal groupings, and some of the dialect names are ethnonyms rather than toponyms in origin. Since the 1944 exile there has been a good deal of dialect mixture and

standardization: a number of high mountain villages were not reinhabited after the exile, and their occupants were resettled in various lowland villages. In addition, the plains dialect has been systematically taught as the literary standard.

Some reshuffling of lowlands and highlands dialects has evidently occurred in the past as well. Chechen linguists and historians speak of a descent of mountain people to the lowlands beginning in about the 13th century, with a later phase beginning about the 16th or 17th century. The standard reason for this descent given in Soviet sources is that the Golden Horde (the Turko-Mongol steppe empire that ruled eastern Ukraine and southern Russia from 1240 until the 16th century) occupied the lowlands and forced the indigenous Caucasians into the mountains, causing a land shortage there; as the Golden Horde weakened, the Caucasians naturally descended to the plains. The real reason was surely the Little Ice Age, a global cooling that began in the 13th century and peaked in the 17th to early 19th centuries, triggering glacial advances and avalanches, lowering the treeline, and shortening the growing season in mountains and therefore making high mountain areas uninhabitable (see Grove 1988, especially pp. 200ff.). Central Caucasian upland villages and pastures must have been abandoned as Alpine ones were, and the result would have been dialect mixture and reforming of dialect boundaries.

In addition, climatic conditions changed the economic conditions and consequently the sociolinguistic ranking of highland and lowland dialects: in cooler periods the north Caucasian plain would have received more rainfall and experienced increased carrying capacity, attracting immigrants from the highlands. As the lowland economy prospered and the highland economy faltered, lowlands dialects acquired prestige and their dialect innovations spread. The modern dialects of Ch evidently postdate the 13th century (the beginning of the Little Ice Age) and may well antedate the 17th century (the first major glacial advance in the central Caucasus probably dates to 1640–1680: Grove 1988: 201ff.), but the actual phonological changes that define the dialects include an earlier level and a later one. The split of Batsbi from Chechen-Ingush must antedate the Little Ice Age, and the split of Ingush from Ch must have approximately coincided with the Little Ice Age. Batsbi, Ingush, and all but the easternmost dialect of Ch have vocalic umlaut of some kind, though the conditions and effects differ; the tendency for umlaut must therefore be a southwestern (i.e. western highland) dialectal one in Proto-Nakh, and the Ch dialects with minimal umlaut and the one with-

out it were on the Proto-Nakh dialect periphery even before the Little Ice Age began. Similarly, posttonic vowel reduction has a western and probably southwestern origin and is Proto-Chechen-Ingush in age. On the other hand, the conditions and effects of umlaut that differentiate the Ch dialects are obviously more recent; the development of front rounded vowels is a recent development, limited to central and eastern lowlands dialects, and evidently reflects Turkic influence. Differences in consonant phonotactics may also be relatively recent. Archaisms such as minimal or no umlaut and retention of Proto-Nakh postvocalic voiced stops are found in the eastern dialects, marking them as the ancient periphery. Umlaut is likely to have arisen in the western highlands, to judge from its present distribution, but at present the lowlands implementation of umlaut (and chiefly that of the central and eastern lowlands) is the most far-reaching and also the most influential. I interpret these various facts as indicating that the split of Batsbi from Chechen-Ingush occurred shortly before the beginning of the Little Ice Age, the split of Ch from Ingush approximately coincided with the beginning of the Little Ice Age, and both that split and the subsequent history of Ch dialects reflect a shift of the Chechen-Ingush sociolinguistic center from the western highlands to the eastern lowlands.

The major dialects of Ch are:

Äqqi (in R, *akkinskij*): the easternmost lowlands dialect, spoken in Daghestan. The Äqqi preserve an oral tradition of relatedness to the La:m-Aqqi ('mountain Äqqi') of the western mountains. The Äqqi dialect resembles western ones in its posttonic neutralization and some morphological features, but the surface phonology and phonetics of its umlaut are features of the central and eastern lowlands where it is now spoken.

Č'a:birloj (in R, *čeberloevskij*), the easternmost mountain dialect. The most archaic dialect, completely lacking umlaut and posttonic neutralization.

Šaroj (R *šaroevskij*), evidently spoken in the central to eastern foothills (no source known to me gives its location).

Itum-Qalle (standard *Iton-Qälla*, R *itum-kalinskij*), a dialect of the central mountains.

Galanč'ož (R *galanč'ožskij*), the western dialect, comprising four varieties: Našxoj, Lam-Aqqi, T'erloj, Melxi).

Kisti (Ch *Kisting*, R *kistinskij*), spoken in Georgia in the upper Alazani and Andi Kojsu basins, though phonologically showing affinities to the western Ch dialects. The Kisti speakers I have heard give evidence of Georgian

influence on their phonetics (and are bilingual in Georgian), though this may be a recent phenomenon.

Ingush is spoken to the west of the westernmost Ch dialects, from the foothills to the high mountains. Batsbi is spoken in Georgia, in the foothills of the southern Caucasus slope.

47.4.2. *Dialect differences in consonants*

The consonant systems of the various dialects of Ch are virtually identical in both phonetics and phonology. There are some phonetic differences as noted above: the voiced pharyngeal segment, ordinarily an epiglottalized glottal stop, is a spirant word-initially in the Pharčxoј variety of the Äqqi dialect (Arsaxanov 1969: 33–34); the geminate stop and affricate series, which in most dialects have little or no aspiration and are sometimes preaspirated, are aspirated in the Kisti dialect (or at least in the pronunciation of the two Kisti speakers I have heard).

Imnaišvili 1977: 275 mentions sporadic retention of Proto-Nakh post-vocalic voiced consonants in dialects; the normal reflex is a glide or zero. The following examples of these retentions are from Arsaxanov 1969: 23, 34. The last Äqqi form is from the Pharčxoј variety. Batsbi forms, which are conservative in their consonantism, are given for comparison where I could find them.

Äqqi	Plains	Batsbi	
<i>küda</i> (p. 23)	<i>kuj</i>	<i>kud</i>	‘cap’
<i>moda</i> (p. 23)	<i>muo</i>		‘scar’
<i>ʃabag</i> (p. 34)	<i>ʃajg</i>	<i>ʃabik’</i>	‘spoon’

Postvocalic voiced obstruents occur frequently in all Ch dialects, but normally they reflect Proto-Nakh ejectives (e.g. Ch *dwog*: Batsbi *dok* ‘heart’; Ch *mwoz/mwož*: Batsbi *moc* ‘honey’).

Proto-Chechen-Ingush *ʒ, *ʒ generally yield voiced fricatives *z*, *ʒ* in most dialects. I have heard *ʒ* in the plains dialect only in initial position, and only in some words with Proto-Chechen-Ingush affricates. (In Ingush, affricates are regular in final position.) Imnaišvili 1977: 262ff. shows affricates retained regularly only in the Xildeħaroј dialect of the high eastern mountains.

If the segmental inventory of consonants differs little from dialect to dialect, there are clearer differences in phonotactics and phoneme distribution. A standard characteristic for dividing Ch dialects is the treatment of initial **st*-, which is variously *st*-, *s*-, and *t*- in different dialects. Here and below, the plains dialect is the conservative one.

Plains	<i>stag</i>	
Č'a:birloj	<i>stag</i>	(Imnaišvili 267)
Itum-Qalle	<i>sag</i>	(Arsaxanov 90; Imnaišvili 267)
Šaroj	<i>sag</i>	(Arsaxanov 86; Imnaišvili 267)
Äqqi	<i>sag</i>	(Arsaxanov 26)
Kisti	<i>sag</i>	(Arsaxanov 156)
western	<i>sag</i>	(Arsaxanov 114, 115, 121, 138)
cf. Ingush	<i>sag</i>	
Batsbi	<i>st'ak'</i>	

In addition, I have heard the pronunciation *tag* in lowlands speech.

Treatments of affricates and fricatives adjacent to other consonants also differentiates dialects. Some dialects turn an affricate into a fricative before the *-n-* of the converb ending **(i)ne*:

Verb stem	Converb	Plains	Šaroj	western (Našxoj)	Vedeno
<i>la:c-</i>	<i>la:c-ina</i>	<i>läcna</i>	<i>lasna</i>	<i>läsna</i>	<i>läsna</i>

Some dialects turn *s* and *š* into affricates after sonorants:

	<i>*urs</i> 'knife'	<i>*nars</i> 'cucumber'
Plains	<i>urs</i>	<i>närs</i> (umlaut in this form unexplained)
Šaroj	<i>urc</i>	<i>narc</i>
Kisti	<i>urc</i>	
cf. Ingush	<i>urs</i>	<i>nars</i>

Others turn affricates into fricatives in the same position:

	<i>*buorc</i> 'millet'	<i>*xa:rc</i> 'wrong'	<i>*burč</i> 'pepper'
Plains	<i>buorc</i>	<i>xa:rc</i>	<i>burč</i>
Itum-Q'ale	<i>buors</i>	<i>xa:rs</i>	<i>burš</i>
western: Našxoj	<i>buorc</i>	<i>xa:rca</i> ⁿ 'overturn'	
cf. Ingush	<i>buorc</i>	<i>xa:rc</i>	<i>burč</i>

As discussed in § 47.5.2, in root-final clusters of the shape **-RC* the pre-consonantal sonorant is a generic (or neutralized, or underspecified) sonorant whose realization as *l*, *r*, or *x* is determined by the following consonant. The Äqqi dialect has *r* before *x* where the other dialects have *l*:

Plains	<i>bwolx</i>	<i>ma:lx</i>
Äqqi	<i>bworx</i> 'work'	<i>ma:rx</i> 'sun'
cf. Ingush	<i>balx</i>	<i>ma:lx</i>
Batsbi	<i>botx</i>	<i>matx</i>

The Itum-Q'ale and Šaroj dialects have *r* before *k* where the other dialects have *x* (Imnaišvili 1977: 268–69; Arsaxanov 1969: 89).

	Plains et al.	<i>mwoxk</i> ‘land’	<i>daxka</i> ‘mouse’
	Itum-Q’ale	<i>mwork</i>	<i>darka</i>
	Šaroj	<i>mwork</i>	<i>darka</i>
cf.	Ingush	<i>mwoxk</i>	<i>daxka</i>
	Batsbi	<i>moxk’</i>	<i>daxk’o</i>

47.4.3. *Dialect differences in vowels*

Vowel systems differ markedly among the dialects, but the differences are almost entirely due to differential operation and effects of the secondary assimilatory processes of umlaut (described for the standard language in § 47.3.1). All dialects have the basic five-vowel underlying inventory (Table 47-5).

Table 47-5. Chechen Dialect Vowels

i	u	i:	u:
ye	wo	ie	uo
a		a:	

Diphthongization of the mid vowels is apparently pan-Ch and can be reconstructed for Proto-Chechen-Ingush. (Imnaišvili 1977: 22–23 describes the mid vowels of the Pharčxoĵ variety of the Äqqi dialect as undiphthongized, but regards this as a secondary development. Arsaxanov 1969 mentions diphthongized *wo*, *uo* and by not discussing the quality of **e*, **e*: gives the impression that diphthongized *ye*, *ie* are also found there.) The Ingush mid vowels are also diphthongized; those of Batsbi are not. Diphthongization evidently applied only to tonic syllables, as the only dialect in which post-tonic vowels are not neutralized to /a/, the Č’a:birloĵ dialect, has undiphthongized short [e] and [o] in posttonic syllables.

The full roster of vocalic assimilatory processes, in approximately the order supported by internal reconstruction and cross-dialectal comparison, is the following. The symbol “>” shows reconstructable ordering and synchronic ranking in implicational hierarchies.

Umlaut

- rounding > fronting
- affecting low > nonlow vowels
- before high > nonhigh vowels

Raising (of mid vowel before high vowel)

Neutralization of posttonic short vowels

mid vowels become high

all vowels become *a*

The dialects of Ch differ in the number of these processes that apply, but apparently not in their ordering. Another notable difference is the presence vs. absence of front rounded vowel phonemes /ü, ü:, wö, üö/: these phonemes occur in central and eastern lowlands dialects, where they are produced by fronting of back vowels and rounding of front vowels (as described for the standard dialect in § 47.3.1), while in the western dialects fronting of back vowels produces /i, i:, ye, ie/ and rounding of front vowels either does not occur or produces /u, u:, wo, uo/.

The following examples from dialects illustrate some of these differences for key dialects. Forms are taken from Arsaxanov 1969 and Imnaišvili 1977 (where more examples and coverage of more dialects can be found). For simplicity examples are limited to verbs with three different root vowels: **a*:, **o*:/uo/, and **e*:/ie/ (for other root vowels, and examples taken from noun inflection, see Arsaxanov 1969 and Imnaišvili 1977). Ingush forms are given for comparison. (Ingush lacks the simple past in *-*in*, so I cite the past in *-*ira* for Ingush.) Blanks mean that that particular form is not attested in the dialect sources available to me. Angle brackets mean that I have constructed the form on the firm evidence of identical umlaut in other lexical items and/or explicit statements in the primary sources.

Forms of *tuox*- 'strike' (and, where those are lacking, *duož*- 'fall'):

	Infinitive * <i>tuox-an</i>	Present * <i>tuox-o</i>	Converb * <i>tuox-uš</i>	Simple past * <i>tuox-in</i>	Past verb * <i>tuox-ine</i> ¹
Č'abirloj	<i>tuoxaⁿ</i>	<i>tuoxo</i>	<i>tuoxuš</i>	<i>tuoxiⁿ</i>	<i>tuoxne</i>
Itum-Qalle	<i>tuoxaⁿ</i>	<i>tuoxa</i>	<i>tu:xuš</i>	<i>twi:xiⁿ</i>	<i>twexna</i>
Šaroj	<i>tuoxaⁿ</i>	<i>tuoxo</i>	<i>tu:xuš</i>	<i>tu:xiⁿ</i>	<i>tuoxna</i>
Äqqi	<i>tuoxaⁿ</i>	<i>tu:xu</i>	<i>du:žuš</i>	<i>tü:xaⁿ</i>	
Plains	<i>tuoxaⁿ</i>	<i>tu:xu</i>	<i>tu:xuš</i>	<i>tü:xiⁿ</i>	<i>tüöxna</i>
Ingush	<i>tuoxa</i>	<i>tuox</i>	<i>tuoxaž</i>	<i>tiexar</i>	<i>tī:xa:</i>

Forms of *D=a:x*- 'live', a verb with present tense in *-*e*; or *a:l*- 'say' (cited only in the past, as its present is in *-*o*).

1. The Č'abirloj past verb form suggests a suffix *-*ine*. The other comparative and internal evidence indicates *-*ina*.

	Infinitive <i>*d=a:x-an</i>	Present <i>*d=a:x-e</i>	Past <i>*d=a:x-in</i> or <i>*d=a:x-ira</i>
Č'a:birloj	<i>da:xaⁿ</i>	<i>da:xe</i>	<i>a:liⁿ</i>
Itum-Qalle	<i>da:xaⁿ</i>		<i>e:liⁿ</i>
Šaroj	<i>da:xaⁿ</i>		<i>e:liⁿ</i>
Äqqi	<i>da:xaⁿ</i>	<i>dä:xa</i>	< <i>dä:xara</i> >
Plains	<i>da:xaⁿ</i>	<i>de:xa</i>	<i>de:xira</i>
Ingush	<i>da:xa</i>	<i>da:x</i>	<i>ä:lar</i>

Forms of *D=ieš-* 'read':

	<i>*d=ieš-an</i>	<i>*d=ieš-o</i>	<i>*d=ieš-uš</i>
Č'a:birloj	<i>diešaⁿ</i>	<i>diešo</i>	<i>diešuš</i>
Šaroj	<i>diešaⁿ</i>	<i>diešo</i>	<i>di:šuš</i>
Itum-Qalle	<i>diešaⁿ</i>	<i>dieša</i>	<i>diešuš</i>
Äqqi (Pxarčxoj)	<i>diešaⁿ</i>	<i>dü:ša</i>	< <i>dü:šaš</i> >
Plains	<i>diešaⁿ</i>	<i>düöšu</i>	<i>düöšuš</i>
Ingush	<i>dieša</i>	<i>dieš</i>	<i>diešaž</i>

Table 47-6 shows which processes are present in which dialects. Umlaut can be broken down into four processes—rounding of **a* vowels, fronting of **a* vowels, fronting of **o*, **u* vowels, rounding of **e*, **i* vowels—and the number of plus signs in an umlaut column shows which of these four processes are present in each dialect. The four processes form an implicational hierarchy, as shown in the note to the table. The Dŕaj subdialect of Č'a:birloj, which is not illustrated above, has rounding but not fronting of low vowels and therefore justifies the ordering of the first two umlaut processes (example: present tense *o:lu* from *a:l-* 'say'). The dialects illustrated above justify the rest of the orderings.

The Č'a:birloj dialect (spoken in the eastern mountains) is the most conservative: it has no umlaut, raising, or posttonic neutralization, and as a result its surface phonemic forms look much like underlying or Proto-Nakh forms. The Šaroj dialect (location not given in any source known to me; probably central foothills) has minimal umlaut as far as affected vowels are concerned: only **a*, **a:* undergo umlaut. The Itum-Qalle dialect (of the central mountains) has minimal umlaut as far as triggering vowels are concerned: only **i* and **u* trigger umlaut. The plains dialect (the literary standard) and the Äqqi dialect (spoken in the eastern lowlands) are quite similar, and have extensive umlaut; the main differences are that the Äqqi dialect subsequently neutralizes all short posttonic vowels to /a/, while the plains dialect preserves posttonic /u/ in verb inflection, and that the Pxarčxoj

Table 47-6. Chechen Dialect Umlaut Processes^a

	Umlaut before *i, u	Umlaut before *e, o	Posttonic *e, *o > *i, *u	Raising before *u, i	Posttonic neutraliza- tion to /a/ /ö/, /ü/	Phonemic
Č'a:birloj	–	–	–	–	–	–
Dɣaj	?	+	–	–	–	–
Šaroj	++	++	–	+	–	–
Itum-Qalle	+++	–	–	+	partial	–
Äqqi	++++	++++	+	+	+	+
Plains	++++	++++	+	+	mostly	+
Ingush	+++	+ (/Co only)	–	partial	+	–

- a. + Rounding of *a
 ++ Fronting of *a (in addition to above)
 +++ Fronting of *o and/or *u (in addition to above)
 ++++ Rounding of *e and/or *i (in addition to above)

variant of the Äqqi dialect raises posttonic *e and *o to *i and *u respectively before neutralizing them to /a/ (as shown in the raising of *uo and *ie to /u:/ and /i:/ before *o and *e). The plains dialect also raises posttonic *o to *u (as shown in the present tense verbal ending -u), but only after vowel raising has occurred, since this -u does not trigger raising (cf. pres. *tuoxu* 'strike(s)' < **tuoxo*). Ingush (the closely related western neighbor of Ch) has full umlaut and is much like a lowlands Ch dialect that has unrounded previously front rounded vowels. (Mal'sagov 1925 shows an Ingush front rounded /üö/ resulting from fronting of *uo, but not from rounding of *ie.) The westernmost Ch dialects on the whole resemble Ingush in their umlaut systems.

The Kisti dialect, spoken in the southern Caucasus foothills in Georgia, resembles the western dialects in the phonology of its umlaut system, but is notable for morphologizing umlaut in the verbal present tense system.

	Stem	Infinitive	1–2 person	3 person
Kisti	a:l- 'say'	a:la	ä:l	o:l
Plains	a:l-	a:lan	o:lu	o:lu

Person inflection is otherwise unknown to Ch, whose verbs agree only in gender with the S/O (i.e. in an ergative pattern; these are consistently

ergative languages). The verb ‘say’ shown here, and the others given by Arsanxanov (1969: 163–64), are transitives whose agreement is with the A, i.e. probably on the accusative pattern.

These Kisti verb forms have umlaut in the stem but a zero ending. Zero present-tense endings are found on some verbs in some dialects (e.g. Č’a:birloj present tense *a:l* or *a:lo* from *a:l-* ‘say’; *da:g* or *da:go* from *da:g-* ‘burn (intrans.)’). In Ingush, the regular present-tense ending is zero, but the present tense shows regular umlaut that requires reconstruction of the ending **-o*, and the obvious solution for either phonological description or internal reconstruction is to posit present tense endings (or at least the **-o* ending), let them trigger umlaut, and then remove the ending. (After removal of the ending, closed syllables are shortened.) Historically, however, the Ingush present tense ending is likely to have resulted from a morphological expansion of the zero ending, with consequent morphologization of umlaut, rather than from phonological deletion of final vowels limited to the present tense. The same deletion of endings after umlaut would account for the umlaut and zero endings of the Kisti verbs, although in Kisti the morphology would provide present tense forms {*a:l-e*} (first and second persons) and {*a:l-o*} (third person).

47.5. Distribution and phonotactics

47.5.1. *Positional restrictions*

There are two phonotactic positions relevant for describing the distribution of consonant manners of articulation: initial and non-initial (i.e. post-vocalic). Ch has four distinct manners of obstruent articulation, but only three appear in any given position. This statement is absolute for initial position, but there are a few exceptions for non-initial position. The possibilities are:

Initial	Non-initial
Voiceless aspirate	Voiceless aspirate
Ejective	(Ejective, with restrictions)
Voiced	Voiced
	Geminate
Sonorant	Sonorant

All four obstruent types appeared in both positions in Proto–Northeast Caucasian, but in Nakh initial geminates have merged with plain voiceless aspirates, and in Chechen–Ingush non-initial ejectives have undergone lenition

and appear as voiced stops or affricates (see § 47.7.2). There are a few exceptions to this rule (see § 47.4.2), and they appear mostly in geographically peripheral dialects. In the standard language only one verb root has a non-initial ejective (other than a uvular). It appears in the following two stems:

D=a:t' 'split, break, tear, crack' (intransitive)

iet' 'tear (clothing), split (wood)' (intransitive) (multiple action)

(There are also a few nouns with this exceptional structure, but I have not done a systematic root survey for nouns.)

47.5.2. Clusters

The following consonant clusters are found in indubitably unanalyzable roots (verb roots, and monosyllabic or disyllabic nouns and adjectives with Daghestanian cognates) in Ch. Some occur only in root-final position, some only initially, and some in both positions.

Initial and non-initial:

<i>st</i>	<i>stag</i>	'person'
	<i>üsta</i>	'ram (over a year old)'
<i>t'q'</i>	<i>t'q'a</i>	'20'
	<i>šat'q'a</i>	'weasel'

Non-initial only:

-xC	<i>čaxč-an</i>	'get burned, scorched'
-rC (C = dental, alveolar, or palatal stop or affricate; and, perhaps for adjectives only, any velar or uvular)	<i>bworč</i>	'millet'
	<i>bworz</i>	'wolf'
	<i>byerd</i>	'(river) bank, bluff'
	<i>byerč</i>	'callus, corn'
-IC (C = fricative)	# <i>bełš</i>	'shoulder'
	<i>bwolx</i>	'work'
-CQ (C = obstruent; Q = uvular of same manner of articulation)	<i>berc'q'aš</i>	'clothes'
	<i>lač'q'an</i>	'hide' (intransitive)
	(note also <i>t'q'</i> , initial and final, above)	
-rst	<i>D-erst-ana</i>	'fat' (adj.)

Initial only:

<i>px-</i>	<i>pxiʔ</i>	'5'
<i>tx-</i>	<i>txwo</i>	'we (exclusive)'

Other consonant clusters are produced by inflection and word formation. Loss of **-i-* in the converb ending {-ina} produces clusters such as *-cn-* and *-sn-* in addition to geminates *-ll-* and *-nn-*. A suffix-initial **-i-* is also lost, under similar conditions, from the diminutive suffix **-ig*. Voicing assimilation then takes place, affecting the **-g* and caused by a preceding obstruent:

<i>r-g</i>	<i>bšärg < *bšar-ig</i>	‘eye’
<i>*s-g > s-k</i>	<i>t’üs-k < *t’us-ig</i>	‘occipital bone’

47.5.3. *Geminates*

The fortis series /pp/, /tt/, /xk/, /qq/, /ss/ are phonetically geminate (or, in the case of /xk/, a cluster) but pattern as unit phonemes in the root structure canon (§ 47.1.2.1). They reflect inherited Proto-Northeast-Caucasian fortis consonants, which are unit consonants in the Daghestanian sister languages. Other sources of surface geminates are morphophonemic processes in clusters over morpheme boundaries:

M’rph’ph’nemic	Phonemic	Phonetic	
<i>a:l-in-a</i>	<i>ʔä:lla</i>	[ʔällə]	‘having said’
<i>e:c-in-a</i>	<i>ʔyecna</i>	[ʔyecnə]	‘having bought’
	<i>ʔyecca</i>	[ʔyecçə]	‘id.’ (allegro pron.)

and focus gemination (see § 47.3.7), which can apply to any postvocalic root consonant. The phonetic outputs of these two processes of gemination are apparently identical to the phonetics of the independent fortis consonants (I have not investigated whether focus gemination of /k/ produces [xk], which is the phonetic value of independent {kk}, or [kk] as might be expected if focus gemination is a purely phonetic process).

47.5.4. *Clusters with pharyngeals, or pharyngealization*

Sequences of consonant plus pharyngeal consonant of homorganic voicing occur word-initially in roots. Some examples of near-minimal pairs:

Plain		Pharyngealized	
<i>da:</i>	‘father’	<i>dša</i>	‘there’
<i>pis</i>	‘stingy’	<i>pʰid</i>	‘frog’
<i>žowhar</i>	‘pearl’	<i>žšow</i>	‘hammer’

In principle, these sequences could be analyzed as clusters, as pharyngealization of consonants, or as pharyngealization of vowels; the first analysis is segmental, the latter two prosodic. The second analysis—(prosodic) pharyngealization of consonants—is adopted here because the severe constraints on clusters, especially initial clusters, in Ch disfavor the first (cluster) analysis and because the distributional constraints on pharyngealization center on

consonants and thus disfavor the third analysis (vowel prosody). (Phonetically, as noted above, pharyngealization is audible both in an almost segment-like acoustically compacted delay in voice onset of the vowel and in distortion of the formant transitions in the following vowel. Contrast Ingush, where the segment-like association with the preceding consonant is often absent and pharyngealization of the vowel seems to be the most salient cue. A revealing comparison is the words for ‘one’: Ch *čaʔ*, Ingush *caʃ*. The two words are almost identical phonetically, but the pharyngealization is centered on the vowel in Ingush and hence pharyngealizes the following consonant, originally a glottal stop.)

The following distributional constraints apply to consonant pharyngealization: (1) Only the more anterior consonants (labial, dental, alveolar, palatal) can be pharyngealized; velars and uvulars cannot. (The independent pharyngeal segments /ʕ/ and /ħ/, as noted above, might be analyzed as pharyngealized glottals /ʔ/ and /h/.) (2) Glottalized consonants cannot be pharyngealized. Ingush has one counterexample, Ch none: Ingush *čʕoaya*, Ch *čoya* ‘firm(ly), strong(ly), very’. (3) Pharyngealization occurs before all vowels, at least in literary Ch, but its lexical frequency is far greater before the low vowels /a/ and /a:/. (4) Consonant pharyngealization occurs in the following parts of speech: nouns, deictic adverbs, numerals, and deictic preverbs (which are usually cognate to deictic adverbs). It does not occur in verbs, adjectives (contrast Ingush, where it does occur in adjectives), or personal pronouns. (5) Consonant pharyngealization is significantly more frequent than expected with nouns of J gender. None of these constraints (with the possible exception of (3)) apply to the pharyngeal segments /ʕ/ and /ħ/, which have few distributional constraints. Thus pharyngealization and pharyngeal segments are phonologically quite distinct (though the acoustics of pharyngealization is identical for both).

47.5.5. *Lexical frequencies; root and syllable canon*

The root form is C*V(C*), where C* = single consonant or one of the permitted clusters and where the initial consonant can be an alternating gender prefix. Many nouns have unanalyzable roots that are longer; evidently a substantial ancient pre-Proto-Nakh suffixal derivational morphology for nouns has frozen to yield longer modern roots, e.g. *patar* ‘shoot, young plant’, *čʕormak* ‘leather strap, including slingshot strap’, *dermat* ‘tanning (of leather)’, of which the last two are probably old compounds—the sequence *-rm-* does not appear in native roots—but are not synchronically analyzable

as far as I know. There are also borrowed words, especially nouns and especially from Arabic and R, with longer roots. Neither borrowing nor ancient suffixal morphology is evident in verbs; there are no borrowed verbs, and the basic root canon $C^*V(C^*)$ is the almost transparently synchronic root form for all verbs. Adjectives resemble verbs in the transparency, regularity, and simplicity of native roots, but they resemble nouns in that borrowings are numerous.

At the morphophonemic level, both long and short vowels can occur in syllables of all types. At the phonetic level, long vowels are shortened in closed syllables (see § 47.3.3).

It is worth considering whether the Ch root canon can be described as accessive or decessive overall (since these terms are occasionally used in historical Caucasian linguistics). Accessive refers to consonant sequences, either clusters or the separate root consonants, which progress from back to front in the mouth; decessive, front to back. A frequency count of consonants in native (or apparently native) roots yields the following results (based on lumping consonants into class by manner and also by point of articulation). In the verb root, in initial position voiceless stops are frequent while glottalized and voiced stops are infrequent; uvulars are frequent while labials and velars are infrequent. In contrast, in the noun root it is the voiced stops that are frequent; labials and dentals are frequent, and pharyngeals infrequent. The high frequencies of labial and dental points and voiced manner of articulation in nouns may (or may not) be due to freezing of gender markers—all of which are anterior consonants and voiced consonants in Nakh as in Daghestanian—to noun initials. Whatever the historical reasons, it is clear that nouns and verbs have very different structural preferences (recall also that consonant pharyngealization is found in nouns but not in verbs).

A better sense of root structure can be gained by comparing initial and final frequencies of various classes of consonants to the frequencies that would be expected (namely, the mean of the initial and final frequencies) if preferences were identical for the two positions. For Ch such a count is feasible only for verbs and adjectives (since noun roots are often extended by apparent fossilized former suffixal morphology, as mentioned above), and I performed a count of the verb roots. In initial position, lateral and uvular points of articulation and the voiceless (aspirated) manner of articulation are significantly more frequent than expected. In final position, alveolar and palatal affricates ($/c/$, $/č/$, etc.) and the voiced manner of articulation (which

reflects the Proto-Nakh and Proto-Nakh-Daghestanian glottalized series) are significantly more frequent than expected; the velar point of articulation is less frequent than expected. Thus there is no evident correlation between initial and final points of articulation that would enable one to describe the verb root as either decessive or accessive overall. But there are clear asymmetries in initial and final preferences. (In the first count described here—high vs. low initial frequencies—the high and low frequencies were defined as one standard deviation or more from the mean frequency for all classes of consonants. In the second count—initial vs. final frequencies in verb roots—significance levels were determined by chi-square test, with probabilities of 0.05 or less deemed significant. Both counts were based on a complete survey of the words in Maciev 1961.)

47.5.6. Text frequencies

Text frequencies of consonants roughly mirror their lexical frequencies, except that the consonants involved in gender prefixation (*b*, *d*, *j*, *v*) have extremely high text frequencies. The text frequencies in Table 47-7 are based on my count of a transcribed recorded text. They show that labial stops have very low frequency (except for *b*, which is a gender marker), but *m* is fairly frequent; otherwise, dental and alveolar consonants have high frequency. Ejectives have low frequency. The gender marks *b*, *d*, *j*, *v* have high frequency among initial consonants, and *n*, *l*, and *r* have high frequency in non-initial position because of their high frequency in inflectional affixes. In general, then, wherever text frequency is not skewed by the shape of high-frequency inflectional morphemes, Ch disfavors labials (other than *m*) but is otherwise like a European language in its generally high text frequencies of more anterior consonants. The “exotic” consonants in the inventory—uvulars, pharyngeals, ejectives—are of low text frequency.

These figures are consistent on the whole with the results of the more limited survey reported in Trubetzkoy 1969: 264–65.

Plain vowels vastly outnumber pharyngealized ones in text frequency. For the following tabulation, a vowel was counted as pharyngealized if it followed *Cʕ* or *Cħ*; vowels following the independent pharyngeal segments are counted separately.

	Root syllables	Other	Total
Plain	224	192	416
Pharyngealized	11	0	11
After /ʕ/, /ħ/	18	2	20

Table 47-7. Chechen Consonant Frequencies

	Initial	Non-initial		Initial	Non-initial
p	0	8	ʒ	0	0
t	19	38	ʒ̥	0	0
c	23	16	g	2	32
č	5	26	s	16	29
k	9	7	š	25	20
q	9	4	x	41	39
ʔ	84 ^a	5	ħ	21	29
ʕ	6	0	h	5	0
p'	0	0 ^b	z	3	30
t'	9	0 ^b	ž	8	3
c'	2	0 ^b	ɣ	3	11
č'	3	0 ^b	m	19	12
k'	5	0 ^b	n	10	53
q'	5	3	l	7	74
b	48 ^c	1	r	0 ^d	105
d	73 ^c	10	v	30 ^c	2
			j	18 ^c	0

a. Initially, same as vowel-initial.

b. Non-initially, phonotactically prohibited.

c. Gender prefix (initial).

d. Initially, phonotactically prohibited.

47.6. Suprasegmentals

In Ch, short vowels are quite short and long vowels quite long. There appears to be no reason, phonetic or morphophonemic, to posit a mora structure. Long vowels are either original, inherited from Proto-Nakh (e.g. *da:q'a* 'portion, share, part', cf. Batsbi *da:q*), or secondary, the result of intervocalic consonant loss (e.g. *be:r* 'child', cf. Batsbi *bader*).

Stress is word-initial. When a verb takes a preverb, both verb and preverb roots are stressed (I judge the preverb stress to be phonetically primary and the verb root stress secondary). The main acoustic correlates of stress

are (1) full inventory of vowels (unstressed syllables undergo posttonic neutralization of short vowels and some phonetic shortening of long vowels); (2) accent peaks in phrasal and sentential prosody.

Three clitics carry inherent high pitch: negative *ca*, negative imperative (and emphatic) *ma*, and coordinating (also scope or focus marker) *ʔa*. The first two are proclitic, the third enclitic. The proclitics have high pitch but do not attract word-initial stress from the following word. Thus (using ' to indicate stress and an acute accent to indicate high pitch):

cá 'vieza 'doesn't like'
má 'tuoxa 'don't hit!'

In many contexts final unstressed short vowels (all of which are /a/ = [ə] as the result of unstressed vowel neutralization) are devoiced or lost. An example is the following, where the enclitic *ʔa* is used to mark one type of clause chaining. Thus we have:

a:xč̣a: ʔa della
 money & having given
 [axč̣a: ʔə dəllə] or [axč̣a:ʔ dell]
 'having given (someone) money, ...', 'gave (someone) money and ...'

Sentence intonation is fairly level on the whole. The overall contour is a falling one for both statements and questions. Yes–no questions, however, are marked with a verbal suffix *-Vj* which has high pitch. In addition, in chained clauses of the type most closely bound to the verb, each clause (chained or final) has a distinctive downstepped intonation contour. In each chained clause, the VP (or similar projection of V) is the downstep domain: the first root syllable is high, the next low, the next high with downstep, and so on. Downstep is reset at clause boundaries, though a declination contour extends over the entire sentence (i.e. over the entire series of chained clauses). Nonfinal chained clauses may end on either a high or low pitch; final (finite) clauses always end on a low pitch. (I do not know whether word order is manipulated to achieve this final low pitch. Final clauses in chains often have subject-final order, which places a VP-external word after the downstep domain and guarantees sentence-final low pitch.) This is only an approximate description; a basic description of downstep is still lacking.

In the following example, taken from a carefully and artistically narrated text, the speaker intones each phrase distinctly and consequently treats two NPs as downstep domains as well (although downstep applies root-internally to syllables in these NPs, while in the VPs it applies to morpheme-initial syllables only). Square brackets mark downstep domains, "H" and "L" in the interlinear mark high and low pitch on syllables, and

“H-” is downstepped high. Phrases not in brackets are not involved in downstep.

[Chana	ʕü:ranna _{NP}]	voqqa	stag	[hala+	yätta-ča _v]	cunna	gwo
	H L H-			H	L H-		
one-OBL	morning-DAT	old	man	up+	stood-CV	he-DAT	see-PRES
[šajn	kertara _{NP}]	[dʃa+	j=üöduš _v]	larš.			
	H L H-	H	L H-	L			
his-RFL	fence-ABL	away+going	tracks				

‘One morning the old man gets up and sees tracks leading away from his fence.’

47.7. Historical phonology

47.7.1. Vowel changes

Umlaut, raising, and posttonic neutralization are the chief historical changes in Ch vocalism; they are described in § 47.4.

47.7.2. Consonantal changes

Few changes have affected consonants in the reconstructable history of Ch or Chechen-Ingush. An important change since Proto-Nakh is lenition of original ejectives and voiced obstruents in postvocalic position: original ejectives become Chechen-Ingush voiced obstruents (except that **q* remains unchanged), and original voiced ones become glides or are lost entirely. These changes do not affect Batsbi, which retains Proto-Nakh postvocalic consonants intact. Since the question of whether ejectives become voiced stops or vice versa has provoked some discussion in both Caucasian and Indo-European studies, and since Ch and Ingush give clear evidence for a shift from ejective to voiced, the evidence is worth laying out in some detail. Table 47-8 presents some cognate sets.

A number of these words have solid Nakh-Daghestanian etymologies complete with regular correspondences and grammatical properties such as gender and valence. The Daghestanian consonant correspondences are well understood (see e.g. Gigineišvili 1977), and the Nakh-Daghestanian consonant correspondences are generally clear and often transparent (see Gigineišvili 1977:21ff., Imnaišvili 1977: 270ff., Nichols 1993). The evidence shows unambiguously that Batsbi consonantism is conservative, and that Chechen-Ingush has undergone a shift from ejective to voiced that is part of a general consonant shift involving lenition. The good cognate sets (with selected cognates) are:

Table 47-8. Proto-Nakh Consonant Cognate Sets

	Proto-Nakh	Chechen	Ingush	Batsbi	
*-b	*ʕabikʼ	ʕajg	ʕajg	ʕabikʼ	‘spoon’ (see §4.2)
*-d	*bader	be:r	bier	bader	‘child’
*-g	*gogo	guo	guo	gogö	‘circle’
*-g	*žagn-	žajn (dial.)	–	žagn, žagno	‘book’
*-g	(loan)	saj	saj	–	Ossetic <i>sag</i> ‘deer’
*-pʼ	*napʼ, nʕapʼ	na:b // nʕa:b	na:b	nʕapʼ	‘sleep’ (noun)
*-tʼ	*šatʼ-	šada // #šodo	šwod	šatʼ	‘pitchfork’
*-cʼ	*D=ucʼ-	D=uz-	D=už-	D=ucʼ-	‘fill’
*-cʼ	(loan)	ärzu:	ärzu:	arcʼiv	Georgian <i>arcʼivi</i> ‘eagle’
*-kʼ	*dokʼ	dwog	dwog	dokʼ	‘heart’
*-qʼ	*D=e:qʼ-	D=ieqʼ-	D=ieqʼ-	D=eqʼ-	‘divide’

**D=ucʼ* ‘fill’: Gigineišvili 1977: 100; cognates: Kibrik & Kodzasov 1990, nos. 70, 71. Some cognates: Andi =*Vcʼi*, Bezhta =*ocʼ*, Lak =*ucʼ*, Dargi (Chirag) =*ircʼ*, Tabassaran *a=cʼ*, Archi =*a:cʼa*, Xinalug *cʼi* (abstract root forms cited from Kibrik & Kodzasov). Proto-NEC **=ucʼ* / *=acʼ* (with possible ablaut) ‘fill’. The reconstructable valence is labile transitive/intransitive, with nominative (absolutive) S/O denoting the vessel filled, oblique ‘water’, and optional ergative agent.

**dokʼ* ‘heart’: Gigineišvili 1977: 82; cognates: Kibrik & Kodzasov 1990: 34 (no. 56). Some cognates: Andi *rokʷo*, Bezhta *rakʷo*, Lak *qqʕukʼ* // *dakʼ* (different sources), Dargi *urkʼi* // *určʼe*, Tabassaran *jukʼu*, Archi *ikwʼ*, Xinalug *ung*. Proto-NEC **rVkʼu* / *Vrkʼu*, gender class 4.

**gogo* ‘circle’ is not attested in that form in Daghestanian, but there are reduplicated cognates meaning ‘round’: Avar *gurgina=*, Hunzib *gerdu*, Lak *kkurkkissa*, Dargi *gulgulze*, Tabassaran *čʼigurgumi* // *gergmi*, Archi *gukkittu=*, Xinalug *gongozly* (Gigineišvili 1977: 78; cognates: Kibrik & Kodzasov 1990: 236 (no. 517). These adjectives find direct parallels in Ch *gworga*, Ingush *gyerga*, Batsbi *gogrʼi* ‘round’—but the medial *-g-* cannot be cognate in these forms. I assume the Batsbi form is conservative and based on the noun **gogo*, while the Chechen-Ingush form has been reinterpreted as based on a noun containing the diminutive suffix **-(i)kʼ*. All three Nakh nouns belong to the same gender class 3.

The Ossetic source for ‘deer’ and the Georgian source for ‘eagle’ (both in Table 47-8) provide good evidence for the earlier consonantism, since neither Ossetic nor Georgian has undergone a recent shift in consonant manners of articulation. (Ossetic is an Iranian language of the central Caucasus, the western neighbor of Ingush. Georgian is a Kartvelian language of the southern Caucasus, the southern neighbor of Ch and Ingush.) These loanwords show that the original consonantism in these words was of the Batsbi type and not of the Chechen-Ingush type.

47.8. Other analyses

Beerle-Moor 1988 proposes an analysis of vowel alternations in verbs which does not use umlaut but rather has stem vowels predict vowels of endings rather than vice versa, and does not treat the vocalism of the infinitive stem as primary. Fallon 1993 also takes this approach. This system does not distinguish diphthongized from undiphthongized mid vowels. It is based on the orthography and appears to generate orthographic forms well, even handling as regular a few verbs whose conjugation is split or irregular in the system used here. That is, it works well synchronically. On the other hand, the system used here, where stem vowels undergo umlaut, is motivated by comparative and historical analysis and can describe the dialectal differences in umlaut discussed in § 47.4.3.

References

- Arsaxanov, I. A. 1969. *Čečenskaja dialektologija*. Groznyj: Čečeno-inguškoe knižnoe izdatel'stvo.
- Beerle, Werner. 1988. “A Contribution to the Morphology of the Simple Verb in Chechen.” In *Studia Caucasologia I*, ed. Fridrik Thordarson, pp. 1–37. Oslo: Institute for Comparative Research in Human Culture, Norwegian University Press.
- Fallon, Paul D. 1993. “Verbal Ablaut in Chechen.” Paper presented at the Eighth Biennial Conference on Non-Slavic Languages [of the Former Soviet Union], University of Chicago, April.
- Grove, Jean M. 1988. *The Little Ice Age*. London: Methuen.
- Imnaišvili, D. S. 1977. *Istoriko-sravnitel'nyj analiz fonetiki naxskix jazykov*. Tbilisi, Georgia: Mecniereba.

- Kingston, John, and Johanna Nichols. 1986. "Pharyngealization in Chechen." Paper presented at the 112th Meeting of the Acoustical Society of America, Anaheim, December.
- Kodzasov, S. V. 1986. "Faringo-laringal'noe suženie v dagestanskix jazykax." In *Aktual'nye problemy dagestansko-naxskogo jazykoznanija*, ed. G. A. Klimov, pp. 16–44. Maxačkala, Russia: Dagfil SSSR.
- Maciev, A. G. 1961. *Čečensko-russkij slovar'*. Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'ny slovar'ej.
- Trubetzkoy, Nikolay S. 1969. *Principles of Phonology*, trans. Christiane A. M. Baltaxe. Berkeley and Los Angeles: University of California Press.

This page intentionally left blank.

Lak Phonology

Gregory D. S. Anderson

University of Chicago

48.1. Introduction

Lak (*lakku maz*), a member of the Northeast Caucasian language family, is one of the five literary languages of Daghestan.¹ It is considered by some investigators to form an individual branch within the family (Bokarev 1981: 6) while others favor a Lak-Dargwa subgroup (Hewitt 1981: 197).² Presently, there are approximately 100,000 speakers of Lak. Investigation of Lak phonology began with Uslar (1890), who worked with a speaker of the Wicxi dialect. Although various other studies devoted to or including Lak data appeared in the interim, e.g. Erckert (1895), Dirr (1928), Trubetzkoy (1931), or Bouda (1949), this remained the standard reference work until Zhirkov's (1955) sketch of Lak grammar. Murkelinski (1971) represents the most recent major Russian-language work dedicated to Lak. Khaidakov (1966) is the standard work on Lak dialectology. Various other studies of specific aspects of its phonology and dialectology have also been addressed in a range of papers, mainly in Georgian and Russian academic journals and publications, e.g. Murkelinski (1949), Cercvadze (1980), or A. Abdullaev (1981). Bokarev (1961, 1981) and Gigineishvili (1977) are the standard Soviet works on comparative Daghestanian (Proto-Northeast Caucasian) phonology. Kibrik and Kodzasov (1990) include a brief phonology of virtually all the Daghestanian languages, including Lak.

1. AUTHOR'S NOTE: Many thanks to Victor Friedman for valuable comments on an earlier draft of this paper. The usual disclaimers apply. ABBREVIATIONS: ABS, absolutive; And, Andi; Arak, Arakul dialect; Arch, Archi; Av, Avar; Axv, Axvax; Balx, Balxar dialect; Darg, Dargwa; DAT, dative; DUR, durative; GEN, genitive; GER, gerund; Lezg, Lezgian; lit., literary (Kumux) dialect; LOC, locative; PDag, Proto-Daghestanian; PL, plural; SG, singular; Tab, Tabassaran.

2. Among the phonological similarities that Lak shares with Dargwa are the development of $*\lambda$ and $*\lambda\lambda$ to $(x)x/(s)s$ and the changes of $*\lambda' > k'$ or $*c'c' > c'$ (only in initial position in Dargwa). Note that Lak and Dargwa share certain morphological features, e.g. the personal inflection system, noun declension, etc. (Gaprindashvili 1954: 322ff.).

48.1.1. Vowels

Following Zhirkov (1955) and Murkelinski (1967), the vowels of Lak are found in Table 48-1, and some minimal pairs follow.

Table 48-1. Lak Vowels

i		u		i ^ɕ		u ^ɕ
	a				a ^ɕ	
<i>č'un</i>	'time'		vs. <i>č'u^ɕn</i>		'rosary'	
<i>č'ira</i>	'wall'		vs. <i>č'iri</i>		'beard'	
<i>biwtssa</i>	'shining (sun)'		vs. <i>buwtssa</i>		'drunk'	
<i>as</i>	'nobility'		vs. <i>a^ɕs an</i>		'to torture/put to death'	
<i>liwčunu</i>	'having thrown'		vs. <i>li^ɕwčunu</i>		'having run'	

Kibrik and Kodzasov (1990) list a fourth unpharyngealized vowel phoneme [ǎ], which is lower (and more centralized??) than plain [a] and is seen in such words as *zăt* 'thing' or *šänma* '3'.

The pharyngealized vowels range in pronunciation among individual Lak speakers from only slightly to distinctly pharyngealized. The pharyngealized [a^ɕ] and [u^ɕ] are pronounced by all speakers with a fronted articulation; in the case of [i^ɕ] and [u^ɕ], the pharyngealization has the "expected" effect of lowering the vowel articulation; thus, these vowels are phonetically realized as [e], [æ], and [œ] or [e^ɕ] [æ^ɕ], and [œ^ɕ]; according to Catford (1994: 54–55) the [œ^ɕ] vowel is highly centralized, but fronted from [u] (and [a]). Of the pharyngealized vowels [a^ɕ] is most common. In words with pharyngealized vowels, some sounds exhibit palatalized realizations, e.g. velars and [l].

Pharyngealization is considered an autosegment in Lak (Kibrik and Kodzasov 1990), and may spread across most segments in the word, e.g. *ya^ɕhu^ɕ* 'bravery' or *a^ɕr^ɕa^ɕb^ɕa^ɕ* 'cart'; a dental sibilant can block the spread of pharyngealization, e.g. *hi^ɕsilu* 'scissors'. Bouda (1949: 9) lists forms that vary as to the realization of pharyngealization on the various vowels in the word, e.g. *ya^ɕtul* ~ *yat'u^ɕl* ~ *ya^ɕt'u^ɕl* 'red'. Bouda also notes that pharyngealization has a tendency to spread through a word, e.g. *ču^ɕryu^ɕlu^ɕt* 'dross, slag'.

In addition to the vowel phonemes listed in Table 48-1, there are rising and falling diphthongs with the glides [y] and [w].³ Certain words are differentiated solely by the presence or absence of a long vowel; thus, vowel length has a limited phonemic load in (at least some dialects of) Lak, e.g. *maq*

3. According to Murkelinski (1971: 51) there is a tendency for -Vy sequences to be stressed; this may simply reflect the tendency for heavy syllables to be stressed, see below.

'word' vs. *ma:q* 'behind'. Vowel length of the expressive type is also encountered in "emphatic" adjectival forms, for example *qunmassa* 'big' vs. *qun-maassa* 'very big', *laqiissa* 'long' vs. *laqiissa* 'very long', or *bu^hxttulssa* 'high' vs. *bu^hxttuulssa* 'very high'. There is no one opinion on the status of the mutual interaction of stress and vowel length in the linguistic literature on Lak; stress and vowel length are described as mutually dependent by some Lak specialists (Zhirkov, Khaidakov) but only partly dependent by others (Murkelinski 1971: 12). See A. Abdullaev (1981) for a discussion and § 48.3.2, esp. (44) to (47).⁴

Lak /a/ exhibits a range of positional allophony. Preceding the glides __w and __y, /a/ is realized as [o] and [ɛ], respectively. Following uvulars or pharyngeals, /i/ surfaces as [ɪ] or [e], while /u/ likewise lowers to [ʊ] or [o] in the same environment, especially following [h]; presumably, these are phonetically pharyngealized as well. Also, /i^h/ exhibits similar allophony as /i/ following [h], that is, it is generally realized as (pharyngealized) [ɪ]/[e] in this position. According to Murkelinski (1971: 12), /a/ may front to [ä] following an unusual array of sounds, namely *h*, *l*, *d*, *k*, *č*, or *c*. According to Uslar (1890) final /-u/ is realized as [o] for some (Wicxi) speakers of Lak, e.g. *arco* 'silver' (= *arcu*) or *nino* 'mother' (= *ninu*).

According to the data of Burzhunov (1984), the status of Russian loans in Lak can be characterized by three gradual stages. The usual sociolinguistic factors of age, place of birth, residence, education, etc. result in a continuum of variability in loan pronunciation from total assimilation to codeswitching in Lak speakers. Pre-revolutionary loans show little phonological predictability with respect to the modifications into the phonological system, for example, one finds such forms as *yanaral* ~ *yiniral* ~ *yinaral*, or even *yananar* for various older regional variants of Russian *general*. The second stage exhibits a relative consistency in loan phonology. The third stage results from mass (/universal) bilingualism and a process akin to decreolization phenomena, with the loans taking on more and more the shape of the Russian source, subject to individual experience with Russian; the final consequence represents a type of codeswitching phenomena in urban bilingual Lak speakers. At stage two, Russian vowels assimilate in the following manner in Lak: [u] and [i] and often [a] are preserved as such. (Stressed) [a] and [u] following a palatal(ized) sound in Russian surface as [a^h] and [u^h], respectively, "soft" [o] similarly becomes [u^h]. (Stressed) Russian [o] gives [u], and Russian [e] and [i] > [i] (1).

4. The frustrating variation of information contained in the various descriptions of Lak is an unfortunate fact of the state of affairs of Daghestanian linguistics in general.

- | | | |
|-----|--|---|
| (1) | <i>turt</i> < <i>tort</i> | <i>bumma</i> < <i>bomba</i> |
| | <i>sawit</i> < <i>sovet</i> ⁷ | <i>lu^hstra</i> < <i>lustra</i> |
| | <i>zara^hdka</i> < <i>zariadka</i> | <i>bukwa</i> < <i>bukva</i> |

Lak exhibits a type of ablaut in a number of morphologically determined environments. This typically involves /a/ and /u/, though alternations involving /i/ are attested. One common environment for this type of ablaut is in the oblique stem of the noun, e.g. the GEN.SG (2). Most examples have an [a]-vocalism in the ABS.SG stem but [u]-vocalism in the GEN.SG stem.⁵ The reverse pattern (i.e. ABS.SG [u] vs. GEN.SG [a]) is attested in several nouns as well. In a few examples, there is an alternation between [a]-vocalism in the ABS.SG following a labialized consonant and [u]-vocalism in the GEN.SG following a plain (non-labialized) consonant.

- | | | | |
|-----|-------------------------|-----------------------------|--------------------|
| (2) | ABS | GEN.SG | gloss |
| | <i>bar^h</i> | <i>bur^hal</i> | 'spine' |
| | <i>ppa</i> | <i>ppur^hul</i> | 'awl' (M '71) |
| | <i>marč^w</i> | <i>murč^wal</i> | 'wind' |
| | <i>ba^hrč</i> | <i>bu^hrčulil</i> | 'calf' |
| | <i>ma^hr</i> | <i>mu^hral</i> | 'nail' |
| | <i>qq^war</i> | <i>qqural</i> | 'bundle (of wood)' |
| | <i>šš^wa</i> | <i>ššuttal</i> | 'line' |
| | <i>c^hu</i> | <i>c^haral</i> | 'fire' |
| | <i>č^wu</i> | <i>č^wal</i> | 'horse' |
| | <i>χχulu</i> | <i>χχalal</i> | 'hay' |

In most ablauting nouns, the ABS.PL has the same vocalism as the ABS.SG, e.g. [a] (I. Abdullaev 1987); however, in some lexical items, there is [a]-

5. Lak nouns have up to four stem patterns, viz. ABS.SG, GEN.SG, ABS.PL, and GEN.PL; all OBLIQUE cases are formed on the basis of the GEN. stem. According to Burčulaje (1984: 134) nominal ablaut in Lak arose from the grammaticalization of phonological changes; specifically, *a* > *u* when there was (i) a labial(ized) ([+Round]) consonant in the word anywhere, or (ii) *u* in a following syllable, i.e. a process of Round Harmony, possibly motivated by some earlier stress system. However, some nouns ablaut where no [+Round] segments are motivated other than to account for the ablaut. The fact that not all words with [+Round] sounds exhibit ablaut might be explained through differing original stress as well. Why the purported Round Harmony process has been grammaticalized the way it has, alternatively affecting Oblique and/or Plural stems, is not understood. As many other Northeast Caucasian languages have ablaut phenomena similar to Lak, it seems likely that there was at the very least a tendency to develop morphological ablaut, if not an actual semi-grammaticalized system in the proto-language itself.

vocalism in all but the ABS.SG., or [u]-vocalism in the GEN.SG stem and in the ABS.PL forms but [a]-vocalism in the GEN.PL.

(3)	ABS.SG	GEN.SG	ABS.PL	GEN.PL	gloss
	<i>ppa</i>	<i>ppurdul</i>	<i>ppardu</i>	<i>ppurdal</i>	'awl' (Kh '62)
	<i>maq</i>	<i>muqqul</i>	<i>maqrū</i>	<i>muqqurt:il</i>	'word'
	<i>ttark'</i>	<i>tturč'al</i>	<i>ttark'ru</i>	<i>tturč'ardil</i>	'bone'
	<i>kac'</i>	<i>kuc'al</i>	<i>kuc'ru</i>	<i>kac'urdil</i>	'bouquet'
	<i>kkark</i>	<i>kkurkkal</i>	<i>kkarkru</i>	<i>kkarurdil</i>	'edge'

Similar ablaut phenomena are attested throughout the Daghestanian languages (Khaidakov 1981), e.g. Avar *č'at'* 'flea' *č'ut'ul* 'PL' or Dargwa *bec'* 'wolf' vs. *buc'ī* 'PL'.

Note that there seems to have been "pharyngealization stability" during ablaut in Lak (4); that is, the pharyngealized articulation associated with the vowel (word?) is maintained when the vowel changes its segmental form as a result of the ablauting process; this process is attested in all the relevant examples.

(4) Pharyngealization Stability	
ɤ	ɤ
σ	σ
/ʌ/	/ʌ/
CVC...	CVC...
[-hi]	> [÷rnd]

On occasion, there are other ablaut patterns attested in Lak; these are highly lexically restricted in the nominal system, e.g. *c^wu* (GEN.SG) *c^wil* 'salt'.⁶

Stress in Lak is not well-understood. Murkelinski notes its connection with vowel length and closed syllables, though open syllables can bear stress. In general, it appears that there is a tendency to stress the last closed or heavy syllable in the root or inflectional affix, e.g. *dúrka* 'food' *buxttan* 'slander' or *búrču* 'skin' vs. GEN.SG *burčúl*. However, certain suffixes attract stress, some of which stress a short (?) vowel in an open syllable, e.g. *-č'í*, *-má*, *-šáwu*, or imperative *-á/-í/-ú*, or case affixes such as *-č'á* 'by', e.g. *č'irač'á* 'by the wall' or *qqalay* 'tin' vs. *qqalayč'í* 'tinsmith'. In the case of *-č'í*, a borrowed affix, the original final stress pattern of the Turkic source may be

6. Such alternations are attested more commonly in other NE Caucasian languages, e.g. Av. *q'ili* 'saddle' *q'olol* 'GEN'. Note the Lak dialectal variation *ssu* 'sister' vs. *ssil* ~ *ssuyil* 'GEN.SG'.

maintained in the Lak forms. Further complicating the issues surrounding stress assignment are so-called “stressless” words, e.g. *tturlu* ‘cloud’ or *ƙatt-araxxucanu* ‘nighttime cowshed’, cf. A. Abdullaev (1981). Thus, although heavy/closed syllables are frequently stressed in Lak, suggesting a quantity-sensitive phonological stress system, stress falls semi-predictably in certain root/affix combinations, suggesting a Lezgian-like morpholexical stress system.⁷ Stress placement and other suprasegmental phenomena in all the NE Caucasian languages, e.g. pharyngealization, so-called “tonal” or “tone” systems, etc., require further investigation.

48.1.2. Consonants

The traditional inventory (Murkelinski 1971: 20) of consonantal phonemes is listed in Table 48-2.

Table 48-2. Lak Consonants

p	t	c	č	k	q	ʔ
p'	t'	c'	č'	k'	q'	
pp	tt	cc	čč	kk	qq	
b	d		(j) ^a	g ^b		
		s	š	x	χ	ħ
		ss	šš	xx	χχ	
		z	ž		ʁ	
m	n					
w			y			
	l, r					

a. Only in Wixli dialect.

b. Often phonetically [ɣ] for many speakers.

Examples of minimal pairs in Lak include (glottalized) *c'un* ‘navel’ vs. *ccun* ‘thicken’ vs. *č'un* ‘time’ vs. *čun* ‘whither’ vs. *ččun* ‘to itch’; *tartssa* ‘rolled up’ vs. *t'art'ssa* ‘worn out’; *ta'y* ‘foal; pack’ vs. *tta'y* ‘trinket’; *k^(w)a* ‘hand’ vs. *ča* ‘whence’; *xama* ‘foam’ vs. *šama* ‘3 (I)’; *ku'al* ‘undersize person.GEN.SG’ vs. *kkut'al* ‘hole.GEN.PL’; *pal* ‘fortune-telling’ vs. *ppal* ‘stable, byre’; *χu* ‘cut (of

7. In Lezgian (Hapelsmath 1993), a morpholexical stress system is found contrasting strong and weak roots and affixes. Perhaps the Lak system reflects traces of a system similar to Lezgian, that has been partly phonologized to a quantity-sensitive stress system.

dress), collar' vs. $\chi\chi u$ 'night' vs. qu 'field'; qqu (*bitan*) 'to smartly dress s.o.' vs. $kkut$ 'hole' vs. $q'ut$ 'knock' vs. $k'ut$ 'malt'; $dazu$ 'border' vs. $bazu$ 'member; seam' vs. $gazu$ 'herd'; gay 'ABS.PL of *ga* deictic' vs. κay 'rumble'; su 'corn bin' vs. ssu 'sister'; $\hbar aw$ 'barking' vs' haw 'grave'.

In addition to the sounds in Table 48-2, some dialects probably have phonemic labialized sounds. The linguistic tradition of Lak does not generally recognize these as individual unit phonemes, even though initial clusters are lacking in the Kumux (literary standard) dialect, and there are a few minimal pairs differentiated by the presence or lack of labialization (5).

(5)	$\check{s}^w a$	'hay-making'	$\check{s} a$	'step'
	$q^w a$	'oath'	$q a$	'wing'
	$c^w an$	'who.DAT'	can	'why'
	$\check{c}\check{c}^w an$	'to love'	$\check{c}\check{c} an$	'foot'
	$c^w u$	'salt'	$c' u$	'fire'
	$k^w u$	'soot'	$k' u \ xun$	'become heavy'

There is a great deal of interdialectal variation between labialized and non-labialized consonants; sometimes one even finds intradialectal (idiolectal) variation within the same word, cf. the $qq^{(w)}at'a$ 'shrub; trap' isoglosses (Murkelinski 1971: 23). Note also that some dialects of Lak, e.g. various *govory* (subdialects) of Wicxi, lack labialized consonants (see § 48.3). In addition, velars spontaneously labialize before $\{[i] \text{ and } [a]\}$ for some speakers (Trubetskoy 1931: 33). Thus, it is clear that no one single account of labialization is possible given the present confusing state of data presentation and analysis in the linguistic literature on Lak. This is just one of many issues in Lak phonology that cannot be resolved without further testing with a range of native speakers from different dialect areas. Note that labialization is also attested phonemically in various other Daghestanian languages, e.g. Avar, Dargwa, Agul, etc., and there are some correspondences in cognate lexemes among the languages with respect to labialized sounds, e.g. Lak $kk^w a\check{c}\check{c}i$ Av. $g^w a\check{z}i$ 'dog'. Thus, although not explicitly reconstructed by Bokarev (1961, 1981) or Giginishvili (1977), Proto Northeast Caucasian may have possessed a series of phonemically labialized obstruents.

Although no clusters are attested word-initially in standard (Kumux) Lak, in some dialects the loss of vowels in initial syllables preceding sonorants has given rise to word-initial clusters (see § 48.3). At morpheme boundaries, clusters of all types arise. Geminate or emphatic consonants (6) occur in word-initial position (or onset-position in general), but become corresponding simplex aspirates when followed by a consonant or word boundary.

Minimal pairs differentiated solely by an emphatic or simplex aspirated consonant are attested, though not that commonly (7).

- | | | | | |
|-----|---------------------------|------------|---------------|-------------------------|
| (6) | <i>ppu</i> | ‘father’ | <i>qqeppa</i> | ‘pinch’ (of salt, etc.) |
| | <i>xxu</i> | ‘night’ | <i>ussu</i> | ‘brother’ |
| | <i>kk^wačči</i> | ‘dog’ | | |
| | <i>laχχan</i> | ‘to dress’ | <i>laχlan</i> | ‘to dress.DUR’ |
| | <i>purššal</i> | ‘fur.GEN’ | <i>purš</i> | ‘fur’ |
| | <i>maqquł</i> | ‘word.GEN’ | <i>maq</i> | ‘word’ |
| (7) | <i>uččin</i> | ‘to bury’ | <i>učin</i> | ‘to say’ |
| | <i>ssuru</i> | ‘stan’ | <i>suru</i> | ‘shroud’ |
| | <i>ššin</i> | ‘water’ | <i>šin</i> | ‘year’ |

Geminate sonorants, namely *-ll-* and *-nn-*, and to a lesser extent, *-mm-*, occur in the literary language mainly due to assimilation of one sonorant to another, e.g. *u’llu* ‘cows’ or *banna* ‘I do’, *šamma* ‘3’. Conversely, dissimilation occurs in some dialects in this environment, e.g. Wicx. *u’ldu* or *banda*; note that the underlying sequence */-rr-/* always dissimilates to *[-rd-]* in all Lak dialects, e.g. *ma^r* ‘nail’ > *ma^rdu* ‘PL’ < *-rr-*. Note however in various inflectional environments, e.g. at the boundary of stem and stem-augment of nouns, assimilation is apparently blocked, e.g. *tur* ‘sword’ vs. GEN.SG *turlil*. Also, as evidenced in § 48.3, *[l]* alternates with *[r]* sporadically in Lak at the *govor* or idiolectal level.

Aside from the distribution of emphatics adduced above, other sounds exhibit some noteworthy phonotactic restrictions. Not only emphatics, but voiced stops are generally lacking in word-final position; the few occurrences are all clearly loans. In general, glottalized obstruents are the most frequent obstruents in word-final position, with the aspirated affricates *[-c]* and *[-č]* being particularly uncommon in this position. In fact, most simplex obstruents in final position morphophonemically alternate with emphatics. Simplex *[x-]* is uncommon in initial position, while initial *[χ-]* occurs mainly in loans; according to Murkelinski (1971), Lak lacks initial ***x^w-* and ***χ^w-*. In final position *[-h]* is rare and *[-m]* is also uncommon; in most instances, it seems to have been replaced in this position by *[-n]*, e.g. *qan* ‘wax’ GEN.SG *qamul*; note the form *t’ama* ‘firewood’ GEN.SG *t’annul* ABS.PL *t’anttu*, where the assimilation to *-n-* may reflect a generalization of */m/* > *[n]* in coda position. Some loans even exhibit this change, e.g. *Iwrahin* ~ *Irwahin* ‘Ibrahim’. Velars and *[l]* tend to have palatalized allophones in pharyngealized words; note the “front(ed)” quality of pharyngealized vowels discussed above. Phe-

nomena such as these led Trubetzkoy (1931) to call the emphatic sounds “emphatic palatalized.” Note that [l] is frequently palatalized in word-final position in the Wicxi dialect of Lak. According to Uslar (1890) *-b- ~ -w- ~ -Ø-* between two [u]’s, e.g. *qqatlubu ~ qqatluu* ‘in the house’.

Lak exhibits a variety of morphologically conditioned consonant alternations, some of which were phonologically transparent diachronically, but have become opaque in the synchronic phonological system. Most typical are the alternations between *x ~ š* and *k ~ č* (equally applying to “emphatics” in both cases, and to glottalized sounds in the latter case). Note that these sets of sounds often stem from the same sources diachronically (see § 48.2 for an outline of Lak historical phonology). The palatal sounds become the corresponding velar ones before consonants (8) in preconsonantal and word-boundary positions.

(8)	<i>šašan</i>	‘to cook/boil’	<i>šaxlan</i>	‘to have boiled/cooked’
	<i>bišin</i>	‘to put’	<i>bixlan</i>	‘to have put’
	<i>cuša</i>	‘bear’	<i>cuxri</i>	‘bears’
	<i>q’unča</i>	‘calf’	<i>q’unkri</i>	‘calves’
	<i>iwč’an</i>	‘to kill’	<i>iwk’lan</i>	‘to have killed’
	<i>ulča</i>	‘land’	<i>ulklul</i>	‘land.GEN’
	<i>burk’</i>	‘pig’	<i>burč’al</i>	‘pig.GEN.SG’
	<i>tark’</i>	‘bone’	<i>tturč’al</i>	‘bone.GEN.SG’; <i>ttark’ru</i> ABS.PL

The *k > č* alternation also occurs before [i] and [a], i.e. synchronically before [–rnd] vowels, but diachronically before [–back] ones, as Lak [a] < PDag *ä; conversely, some forms with *č > k* before [u] in derived forms (9).

(9)	<i>učin</i>	‘to say’	<i>uwkussa</i>	‘said’
	<i>sičča</i>	‘snout’	<i>sikkural</i>	‘snout.GEN.SG’

Another peculiar aspect of Lak morphophonology is the presence of a morphologically conditioned metathesis that affects stems of a particular phonological shape. Namely, with the stem-augments and plural affixes *-ttu* and *-du* (possibly allomorphs), stems ending in liquid + V metathesize to V + liquid (10). Note that sporadic instances of other examples of metathesis are attested throughout the Lak dialects, and among Dagestanian languages in general, cf. I. Abdullaev (1981).⁸

8. Note the form *č’lmu* ‘birdy’ GEN.SG *č’l’mullul*, where an irregular metathesis of *m ~ l* has caused a “regular” metathesis of the liquid and vowel. This form has the absolutive singular *č’l’mlu* in some Lak dialects.

- | | | | |
|------|-----------------------------|------------------|----------|
| (10) | <i>č'atlu</i> 'birch broom' | <i>č'atulttu</i> | 'ABS.PL' |
| | <i>tturlu</i> 'cloud' | <i>tturuldul</i> | 'GEN.SG' |
| | <i>laqru</i> 'stone bench' | <i>laqurttul</i> | 'GEN.SG' |
| but | <i>laqru</i> 'yeast' | <i>laqrul</i> | 'GEN.SG' |

According to Burzhunov (1984), with the same variability factors as for the vowels discussed in § 48.1.1, Lak assimilates consonants in Russian loans in the following manner: plain stops and affricates in Russian frequently surface as emphatic or glottalized ones; continuant + stop initial clusters get a prothetic vowel, generally [i], though sometimes [u]; obstruent + liquid clusters get an epenthetic vowel between the consonants, generally a copy of V_1 in the stem. Russian [v] surfaces as either [w] (~[b]) or [b] (varying idio-/dia-lectally) (11).

- | | | |
|------|----------------------------------|-----------------------------------|
| (11) | <i>ppapka</i> < <i>papka</i> | <i>p'at'run</i> < <i>patron</i> |
| | <i>ttallun</i> < <i>talon</i> | <i>kislat'a</i> < <i>kislata</i> |
| | <i>wiršuk'</i> < <i>veršok</i> | <i>badra</i> < <i>vedro</i> |
| | <i>kkancirt</i> < <i>kancert</i> | <i>ttablicca</i> < <i>tablica</i> |

48.2. Historical phonology

The main works in comparative Daghestanian phonology are Bokarev (1961, 1981) and Giginishvili (1977). All three works outline the phonological histories of the various Northeast Caucasian language groups, viz. Avar-Andi-Dido, Lezgian, Dargwa, Lak, Tabasaran, (and Archi). Proto-Daghestanian was characterized by a fairly expanded consonantal inventory—being closer to that of Avar, rather than Lak—including five lateral continuants (fricatives and affricates), all of which were changed during the history of Lak. In addition, Lak is characterized by devoicing of PDag **b*, **d*, **g* in many words in initial and *intervocalic* position, e.g. *ppa* 'awl', *ttukku* 'donkey' (cf. Darg. *dugu*), or *kk"áčči* 'bitch' (cf. Avar *g"azhi*); similar developments are seen in Lezgian languages, cf. Talibov (1977).

48.2.1. Consonants

The Kumux (or literary standard) dialect preserved the following PDag sounds unchanged: **n*, **l*, **t*, **t'*, **q*, **q'*, **s*, **ž*, **šš*, **c*, **c'*, **č*, and **č'*, as well as **p*, which was rare and (**p'*), which according to Bokarev (1981) may not have existed at all in PDag (12). Kumux is archaic in various respects phonologically; certain Lak dialects (e.g. Pervocovkri) are characterized by the change of palatals to alveolars: (PDag **č*>) PLak > Kumux Lak

[č] : Perv. [c], e.g. *ccan* ‘foot’ (vs. Kumux *ččan*); thus, PDag *č, *čʼ, *ž, and *šš are not preserved in all Lak dialects. This is also true for certain other sounds in various dialects, e.g. (PDag *qʼ >) (P)Lak [qʼ] > Ø/“glottal catch” in Arakul and Balxar, see § 48.3.

(12)	PDag	Lak	gloss	cognate
	*n	<i>nis</i>	‘cheese’	Tab. <i>nis</i>
	*l	<i>ttilikʼ</i>	‘liver’	Darg. <i>dulekʼ</i>
	*p <rare>	<i>par učʼin</i>	‘to flash’	Av. <i>parχχize</i>
	*t	<i>itan</i>	‘to abandon’	Lezg. <i>tun</i>
	*tʼ	<i>du-tʼin</i>	‘to pour’	And. <i>tʼinni</i>
	*qʼ	<i>qʼaqʼari</i>	‘throat’	Arch. <i>qʼaqʼara</i>
	*c	<i>ca</i>	‘one’	Av. <i>co</i>
	*cʼ	<i>acʼ</i>	‘ten’	Tsez <i>ocʼi</i>
	*č	<i>baʼrč</i>	‘calf’	Av. <i>beče</i>
	*čʼ	<i>určʼ</i>	‘9’	Darg. <i>určʼ</i>
	*s	<i>sus</i>	‘rye’	Darg. <i>susul</i>
	*ž	<i>žu</i>	‘we’	Av. <i>niž</i>
	*šš	<i>ššašlan</i>	‘to weave’	Av. <i>bassize</i>

Lak lost all five of the lateral continuants of PDag. *λ and *λλ (or *l and *H) surface as alternatively x/š and xx/šš, respectively: the velar in word-final position and preceding [u], and the palatal fricative preceding a palatalizing vowel, *i or *ä > Lak [i] and [a]. The voiced lateral affricate *dʎ gives Lak [l]. The PDag. lateral affricates *λʼ and *λλʼ generally give glottalized [kʼ] (> [čʼ]) in Lak (13).

(13)	*λ	<i>šam</i>	‘3’	Av. <i>lab</i>
		<i>xula</i>	‘pitchfork’	Arch. <i>lol</i>
	*λλ	<i>ššin</i>	‘water’	Av. <i>λλin</i>
	*dʎ	<i>arlu</i>	‘7’	Tab. <i>urgub</i>
	*λʼ	<i>bakʼ</i>	‘head’	Av. <i>betʼer</i> (*λʼ > kʼ)
		<i>čʼali</i>	‘ear of corn’	Axv. <i>λʼara</i> (*λʼλʼ > kʼ, čʼ)
	*λλʼ	<i>rukʼ</i>	‘yoke’	Av. <i>ruλʼλʼ</i>
		<i>čč^(u)an⁹</i>	‘to love’	Arch. <i>λλʼan</i>

9. According to Bokarev (1981: 31), the Lak form < pre-Lak *kʼan. Note however the reflection of the geminate origin of the lateral affricate in the emphatic initial čč- in Lak. This form for many Lak speakers has a labialized emphatic ččʷ.

The PDag glottalized emphatics *c'c' and *č'č' are reflected as corresponding glottalized simplicia, viz. [c'] and [č']; note that labialized c'c'^w gives Lak čč^w - (cf. similar developments in Tabasaran). *kk is realized as xx/šš with the same conditioning as for the lateral continuants above. PDag *čč was rare and *cc gives Lak [ss]. The voiced affricates of PDag, *dz and *dž, give Lak [z] and [č'], respectively (14). Note that although surfacing as [z] and [c'], respectively, Lak preserves the original PDag distinction between *dz and *c'c', which has been neutralized in such otherwise archaic languages as Avar.

(14)	PDag	Lak	gloss	cognate
	*dz	<i>maz</i>	'tongue'	Av. <i>mac'c'</i>
	*dž	<i>mič'</i>	'nettle'	Lezg. <i>midž</i>
	*cc	<i>ussu</i>	'brother'	Av. <i>wacc</i>
		<i>ssi</i>	'anger'	Av. <i>ccin</i>
	*c'c'	<i>c'uku</i>	'goat'	Av. <i>c'c'e</i>
		<i>mučč^{xw}</i>	'corner'	Arch. <i>mec'c'e</i> (Tab. <i>yučč^w</i> < <i>yucc^w</i>)
	*č'č'	<i>i'rči</i>	'right'	And. <i>hanč'č'il</i>

Proto-Daghestanian *x and *χ both give [h].¹⁰ Note that according to Bokarev (1981: 25–26) PDag may have lacked emphatic *xx and *χχ. Among the alveolar and palatal continuants, the following diachronic changes are characteristic of Lak: *z > z/ss, *ss > ss/s and *š > š/s (15), the conditioning for the variants being opaque.

(15)	*z	<i>kkaza</i>	'hoe'	Darg. <i>gaza</i>
		<i>ssaw</i>	'sky'	Av. <i>zob</i>
	*ss	<i>ssa-</i>	'who.OBL'	Av. <i>ssun</i>
		<i>tisin</i>	'to cut'	Av. <i>ssuzye</i>
	*š	<i>tturš</i>	'100'	Darg. <i>darš</i>
		<i>ars</i>	'son'	Darg. <i>urši</i>
	*x	<i>hanna</i>	'seed'	Av. <i>xon</i>
	*χ	<i>maha</i>	'tale'	Av. <i>marha</i>

The voiced stops of PDag underwent several changes during their development to Lak. *b surfaces as either [b] or [pp], although in some lexically

10. Note that certain Lak speakers show a variation between [x] and [h], especially in onset-position, e.g. *xunin* ~ *hunin* 'to meet; towards'. This variation was attested even in the early Lak sources (Abdullaev 1992), including "Kasyd about the cloak" *žindral-hibu* (< *-*xibu*) > mod. -*šibu* 'rabies; fury' or "Derbent-name" *xurttu* mod. *hurttu* 'together, jointly', etc.

specified examples, as [-m-];¹¹ **g* generally is realized as [kk] (sometimes > [čč]), while PDag **d* exhibits a range of reflexes, including [n], [tt], and [t] (16).

(16)	PDag	Lak	gloss	cognate
	* <i>b</i>	<i>barc'</i>	'wolf'	Av. <i>bac'</i>
		<i>ppa</i>	'awl'	Darg. <i>bu-reba</i>
	* <i>d</i>	<i>č'it</i>	'bug'	Av. <i>č'et'</i> 'flea'
		<i>ttukku</i>	'donkey'	Arch. <i>dogi</i>
		<i>ššin</i>	'water'	Tab. <i>ššid</i>
	* <i>g</i>	<i>kk'ačči</i>	'bitch'	Av. <i>g'aži</i>
		<i>ččikku</i>	'cuckoo'	Av. <i>gigu</i>

PDag. **k* and **k'* etc. split into [k]/[k'] and [č]/[č'], respectively, the latter variants before historically front vowels (e.g. **i*, **ɛ*), or synchronically before [-rnd] vowels. **q* became its glottalized counterpart [q'] and **ɸ* > [qq]. Proto-Daghestanian velar and uvular affricates give a complex array of reflexes. *(*k*)*kx* > *x*(*x*)/š, *(*k*')*k'x'* > *k'*, **gk* gives [q'], **qqχ* > *χ/χχ*, **qχ* > *q*, **q'χ'* > [q'-] but [-ɸ]. PDag **q'q'χ'* > Lak [k'] (17). Thus, the emphatic glottalized affricates fell together with the glottalized lateral affricates yielding Lak [k'], while the corresponding unglottalized affricates exhibit fricative reflexes; the PDag. short glottalized and unglottalized uvular affricates surfaced as uvular stops (or -*ɸ* in final position for **q'χ'*).

(17)	* <i>k</i>	<i>ku</i>	'ram'	Av. <i>kuy</i>
		<i>čani</i>	'light'	Av. <i>kan</i>
	* <i>k'</i>	<i>rik'</i>	'axe'	Tab. <i>yuk'</i>
		<i>lač'a</i>	'wheat'	Darg. <i>ank'i</i>
	*(<i>k</i>) <i>k'x'</i>	<i>k'ulu</i>	'mouse'	And. <i>hink'x'u</i>
	*(<i>k</i>) <i>kx</i>	<i>nax''</i>	'chaff'	Av. <i>nakxu</i>
	* <i>q</i>	<i>q'u</i>	'20'	Tsez. <i>qu</i>

11. Similar, sporadic and unmotivated **b* > *m* changes occur in various Northeast Caucasian languages and are attested in many Turkic languages, including those of the Daghestanian area: Kumyk, Nogai, and Azeri; however, in the Turkic cases these probably represent a hypercorrective spread of a motivated **b* > [m] change when a nasal follows, e.g. Turkish *ben* but Kumyk/Nogai *men*, Azeri *mān*. An example of such an alternation in Daghestanian languages is lit. Lak *ssuwḥat* vs. Balxar Lak *ssuḥmat*, cf. Dargwa *ssuḥbat*, Avar *ssuḥmat* 'wedding party'. Note the class-III marker -*m*- < *b* following nasals in Lak (though not Wicxi, cf. Kumux *qunmassa* vs. Wicx *qunbassa* or Kumux *šamma* vs. Wicx. *šanba*).

PDag	Lak	gloss	cognate
* <i>ʁ</i>	<i>muq(q)</i>	‘back’	Av. <i>muʁ</i>
	<i>qqʷal</i>	‘udder’	Av. <i>ʁʷari</i>
* <i>gʁ</i>	<i>maqʷ</i>	‘tear’	Av. <i>maʷo</i>
* <i>qqχ</i>	<i>laχ</i>	‘ashes’	Av. <i>raqχu</i>
* <i>qχ</i>	<i>qu</i>	‘field’	Av. <i>χur</i>
* <i>qʰχʷ</i>	<i>barʁ</i>	‘sun’	Av. <i>baqʰχʷ</i>
	<i>qʷini</i>	‘day’	And. <i>qʰχʷinu</i>
* <i>qʰqʰχʷ</i>	<i>qʷurčissa</i>	‘bitter’	Darg. <i>qʷutʰqʷusi</i>
	<i>kʷi</i>	‘winter’	Av. <i>χʰχʷin</i>

PDag **w* generally is realized as Lak [b-] in onset position, but preserved as [-w] in coda position. In addition, [w] may arise from PLak **hʷ*, variously surfacing as [h], [w] (or [Ø]) depending on the dialect, see (31).¹² PDag. **r* > [r] in medial position, [l] or [d] in initial position, and Ø in final position (18); note the previously mentioned *r* ~ *l* and *r* ~ *d* assimilation/dissimilation processes. PDag **m* is frequently realized as [-n] in final (coda?) position in Lak. Although attested as a phoneme within the synchronic phonological systems of many NE Caucasian languages, e.g. Lak or Lezgian, Bokarev does not give any cognate sets and does not list ***y* among his Proto-Daghestanian inventory of phonemes. However, Abdokov (1983) lists the PNEC deictic **ya*, with the following reflexes: Tsez *ye-da* ‘this’, Archi *ya-r/b/t* ‘this’, Chamalal *ya*, And. *že*, dial. Dargwa *iy-* (vs. lit *iš*). Note also the common class marker *y-* that surfaces as such in many modern Daghestanian languages.

(18) * <i>w</i>	<i>burkʷ</i>	‘pig’	Av. <i>boʎon</i>
	<i>ssaw</i>	‘sky’	Lezg. <i>ccaw</i>
	<i>haw</i>	‘grave’	Av <i>xob</i>
* <i>r</i>	<i>qurt</i>	‘pear’	Arch. <i>χert</i>
	<i>qu</i>	‘field’	Av. <i>χur</i>
	<i>laχ</i>	‘ashes’	Av. <i>raqχu</i>

12. According to I. Abdullaev (1992), the common Lak deictic *wa* and its various derived forms, and their reflexes in the various dialects of modern Lak suggest that there was a variation in PLak between **hʷa* and **χʷa*. This is not surprising given the developments of PLak **hʷ*- described in § 48.3, coupled with the fact that Lak [h] derives from PDag **x* (and **χ*)—a tendency, with (hypercorrective ?) lexical residue, that has continued from the earliest attested Lak into the modern dialects; see also footnote 10 above. As both I. Abdullaev (1992) and Burčuladze (1985a) note, oblique forms of the deictic with initial [ʃ-] would be regular reflexes of a form that had a variant in **x*-.

PDag	Lak	gloss	cognate
	<i>dak'</i>	'heart'	Av. <i>rak'</i>
* <i>m</i>	<i>maz</i>	'tongue'	Darg. <i>medz</i>
	<i>sun</i>	'straw'	Av. <i>sum</i>

48.2.2. Vowels

According to Bokarev (1961, 1981) the following changes took place from Proto-Daghestanian to Lak: **ä* > [a] (19), **u* > [u] (20), **e* > [a]¹³ (21), **a* > [a] or [u] after labialized consonants in closed syllables (22), and **i* > [i] or [u], the latter when adjacent to a labial(ized) consonant (23).

- | | | | |
|------|----------------|-----------------|--|
| (19) | Lak | gloss | cognate |
| | <i>bar̥k</i> | 'sun' | Tab. <i>ri̥k</i> ; Darg. <i>ber̥ʃi</i> |
| | <i>maq'</i> | 'tear' | Tab. <i>ni̥kʷ</i> |
| (20) | <i>nuz</i> | 'bride' | Av. <i>nuc'c'a</i> |
| | <i>mur̥c'u</i> | 'corner' | Lezg. <i>murt</i> |
| (21) | <i>maz</i> | 'language' | Lezg. <i>mez</i> |
| | <i>nak'</i> | 'milk' | Lezg. <i>nek'</i> |
| (22) | <i>arcu</i> | 'silver' | Av. <i>ʔarac</i> ; Tab. <i>ars</i> |
| | <i>barz</i> | 'moon' | Darg. <i>badz</i> |
| | <i>sun</i> | 'straw, thatch' | Tab. <i>šʷum</i> ; Lez. <i>sam</i> |
| (23) | <i>mič'</i> | 'nettle' | Av. <i>mič'č'</i> |
| | <i>nis</i> | 'cheese' | Tab. <i>nis</i> |
| | <i>busan</i> | 'to say' | Av. <i>bicine</i> |
| | <i>cuša</i> | 'bear' | Av. <i>cī</i> ; Tab. <i>šʷeʔ</i> |

Pharyngealized vowels are sometimes attested in cognate roots in more than one Daghestanian language, e.g. Darg. *daʔq'*/Lezg. *req(χ)* and Tab. *raʔq'* 'road, path'. The status of pharyngealized vowels in Proto-Daghestanian is currently under investigation.

48.3. Dialectal phonological variation

Aside from the Kumux literary standard, several other dialects are enumerated for Lak by Khaidakov (1966), e.g. Bartxi (Balxar), Wicxi, and Ashti-Wixli. Although certain traits are generally characteristic of one dialect or dialect group (e.g. the [š] : [s] isoglosses), interdialectal borrowing, areal

13. Ostensibly, PDag **e* > Lak [i] in open syllables, but Bokarev (1981) gives no relevant examples; the one he gives, Lak *c'uku* 'she-goat' Avar *c'c'e* Tab. *c'ih*, contradicts his own claim.

diffusion and contact phenomena, both between various Lak dialects as well as Lak with other Daghestanian languages (esp. Shadni Lak and Dargwa) have resulted in a complex dispersal of diagnostic dialectal features, even within a single (ostensibly unified) dialect group. For example, one way in which most dialects of the Asti-Wixli group differ from the Kumux literary standard is by allowing emphatic consonants in coda position (these are restricted to onset position in Kumux); however, the Sumbat *aul* of the Pervocovkri dialect area exhibits the same distribution as the literary language.

48.3.1. Consonants

According to Khaidakov (1966), the so-called Ashti-Wixli group of dialects (Pervocovkri, Kaya-Maššixi, and Wixli) differs from the literary standard in a variety of ways. The series of palatal affricates and fricatives of Kumux and various other dialects (Wixxi/Arakul/Bartxi) corresponds to alveolar ones in Pervocovkri; interestingly, Wixli (with a few lexical exceptions) shows a basic split between plain palatal and labialized alveolar (24). The *aul* of Sumbat has a higher number of lexical exceptions to the palatal > alveolar change than do other dialects: this may simply be due to loans from a neighboring *š*-dialect or from literary Lak, or the result of gradual lexical diffusion of the sound change in process; note the Kumux-like lack of coda-position emphatics in this *govor*:

(24)	lit. Lak	Pervocov.	Wixli	gloss
	<i>kisa</i>	<i>cisa</i>	<i>čisa</i>	‘wallet’
	<i>k’issa</i>	<i>c’issa</i>	<i>č’issa</i>	‘finger’
	<i>k’^wilassa</i>	<i>c’^wilassa</i>	<i>c’^wilassa</i>	‘saddle girth’
	<i>čč’^wan</i>	<i>cc’^wan</i>	<i>cc’^wan</i>	‘to love’
	<i>šš’^wa’pa</i>	<i>ss’^wa’pa</i>	<i>ss’^wa’pa</i>	‘lash’

Alveolar stops have exhibited a tendency to move into palatal affricates in Wixli in the historically palatalizing environments (before **ä* (> [a]) and **i* (>[i])), e.g. lit. *hanтта* Wixli *hančča* ‘day’, alternating idiolectally (i.e. in individual speaker’s pronunciation) with palatalized stops. Thus, we find Kumux *d* : *dž* ~ *d’* and *t*’ : *č*’ ~ *t’*’, etc. (25).¹⁴

(25)	<i>daray</i>	<i>d’aray</i> ~ <i>džaray</i>	‘taffeta’
	<i>da’^wbi</i>	<i>d’a’^wbi</i> ~ <i>dža’^wbi</i>	‘war’
	<i>qat’i</i>	<i>q’at’i</i> ~ <i>q’ač’i</i>	‘wedding’

14. In the Suxiašši *govor* of Wixli forms with [tt-] are preserved before [i] (and to a lesser extent before [a]).

Also, the Ashti-Wixli group of dialects preserves emphatic consonants in coda position (26) (seen also in the Kuli and Arakul dialects); some dialects even preserve them in word-initial clusters resulting from the loss of short unstressed vowels in initial syllables preceding resonants (27).

- | | | | |
|------|-----------------|------------------|------------|
| (26) | lit. | Wixli | gloss |
| | <i>maq</i> | <i>maq</i> | ‘word’ |
| | <i>kkunuk</i> | <i>kkunukk</i> | ‘egg’ |
| | <i>ššaʹrult</i> | <i>ššaʹrultt</i> | ‘sparrows’ |
-
- | | | | |
|------|---------------|--------------|-------------|
| (27) | <i>ttaral</i> | <i>ttral</i> | ‘comb(ing)’ |
| | <i>xxiri</i> | <i>xxri</i> | ‘sea’ |

Among the standard features used as diagnostics for dialectal variation in Lak are the treatment of resonant clusters at morpheme boundaries. In the literary language, as mentioned in § 48.1.2, clusters of /-rl-/ > [-ll-], /-ln- ~ -rn-/ > [-nn-], but /-rr-/ > [-rd-]. Pervocovkri exhibits the same pattern as the literary language, while Wixli shows an unusual [-ld-]/[-rd-], but [-nn-] split.

The Kaya-Maššixi dialect is for the most part like Wixli (e.g. *k* : *č* but *kʷ* : *cʷ*, etc.; emphatics in coda position, et al.). However, final long [uu] in Kumux following a resonant is shortened, while final [aa] remains intact : lit. *huluu* vs. Kaya *hulu* ‘lentil’ but lit./Kaya *ššamaa* ‘gravel’ (see § 48.3.2).¹⁵ Some lexical exceptions to the *k* : *č* change are attested, especially in the Maššixi variant, e.g. *kʹissa* ‘finger’.

A typical feature of the Balxar (Bartxi) dialect group is the loss of Lak **qʹ*¹⁶ and concomitant lowering of /u/ > [o] (28); in coda position **qʹ* is generally realized as a glottal catch. Kumux [w] often corresponds to Balxar (and most other dialects) [b], except orthographic /v/ in postconsonantal position when representing the labialized consonants (29).

- | | | | |
|------|------------------|--------------|----------|
| (28) | lit. | Balx. | gloss |
| | <i>qʹuqin</i> | <i>oin</i> | ‘to cut’ |
| | <i>qʹunqʹula</i> | <i>onola</i> | ‘lock’ |
| | <i>qʹanqʹ</i> | <i>anʹ</i> | ‘smell’ |
-
- | | | | |
|------|---------------|---------------|------------|
| (29) | <i>cawa</i> | <i>caba</i> | ‘1’ |
| | <i>xxuway</i> | <i>xxubay</i> | ‘at night’ |

Another feature of Balxar is the reflex of PDag **λʹ* as [tʰ] (as in Avar), rather than [kʰ] as in most other Lak dialects. In addition, Balxar generally shows [ɸ] where Kumux has [g] (30). Balxar has a number of words with labialized

15. In the Gamiyašši *govor* of Kumux a similar phenomenon is attested.

16. Some words with [qʹ] are preserved in Balxar, e.g. *qʹačʹa* ‘wineskin’, see note 18.

consonants that lack such a consonant in Kumux. Also, hypothetical PLak **h^w*- (Khaidakov 1966: 15, 36) is realized as Balxar [w] or Ø (the latter generally before [i]); this PLak sound gives literary [w-] but Arakul [h-] (31). Balxar is an assimilating dialect; thus we find [-ll-] and [-nn-].

(30)		lit.	Balxar (/Xuli)		gloss
		<i>huk'u</i>	<i>χut'u</i>		'donkey foal'
		<i>qquṛk'</i>	<i>qqurt'</i>		'clucking'
		<i>gičča</i>	<i>kičča</i>		'from there'
		<i>k'urglu</i>	<i>k'urκlu</i>		'clay dish'
(31)	PLak	lit.	Balxar	Arakul	gloss
	<i>*h^wilagan</i>	<i>wilagan</i>	<i>ilawan</i>	<i>hilagan</i>	'to rinse out'
	<i>*h^wit</i>	<i>wit</i>	<i>wit</i>	<i>hit</i>	'whey'

Like Wixli et al., Balxar loses unstressed short vowels in initial syllables before resonants; however, unlike Wixli, degemination is operative in relevant Balxar forms (32). Possibly significant is the observation that in these forms, the vowel length of the literary form is frequently reversed.¹⁷ Finally, Balxar exhibits a variety of sporadic vowel correspondences with the Kumux literary standard.

(32)	lit.	Balxar	gloss
	<i>χχaraa</i>	<i>χra</i>	'stem'
	<i>ttaral</i>	<i>tral</i>	'comb(ing)'
	<i>k'ili</i>	<i>k'lii</i>	'saddle'

In addition to loss of **q'* in Arakul (as in Balxar),¹⁸ several other diagnostic features are typical of this Lak dialect. Hypothetical PLak **h^w* gives Arakul [h-] (see (31)). Furthermore, no degemination phenomena in coda position is attested in Arakul (33).

(33)	lit.	Arakul	gloss
	<i>q'uwa</i>	<i>oba</i>	'20'
	<i>χalq'</i>	<i>χal'</i>	'people'
	<i>c'aq'</i>	<i>c'a'</i>	'strength'

17. For information on vowel correspondences and issues in the problem of vowel length in the Lak dialects, see § 48.3.2.

18. In both Balxar and Arakul there are lexical exceptions where **q'* is preserved. This could be nothing more than borrowings from a dialect that preserves this sound, e.g. the Kumux literary standard, or from another language with modification into Lak phonology, e.g. Arakul *q'uwat* 'strength', or they may represent the lexical residue from a lexically diffused sound change.

lit.	Arakul	gloss
<i>ccac</i>	<i>ccacc</i>	'dogrose'
<i>maq</i>	<i>maqq</i>	'word'
<i>č'u</i>	<i>č'u</i>	'voice'
<i>č'uu</i>	<i>č'u</i>	'colostrum'

The Wicxi dialect of Lak is a mixture of *govory* which exhibit a good deal of variation intra-dialectally. Typically, Wicxi is thought to lack labialized consonants, though not all *govory* of the Wicxi dialect do—perhaps there is some interference from knowledge of the literary standard which possesses such sounds, or perhaps these *govory* have undergone a process of Kumuxization, etc. Thus, although described by Khaidakov as a dissimilating dialect, e.g. [-ld-] and [-nd-], some *govory* show [-ll-] and [-nn-]. Kumux long vowels generally correspond to Wicxi short, except in certain *govory*. Some variants of the Wicxi dialect (e.g. Kuma) show the loss of unstressed short vowels in initial syllables preceding resonants, others do not (e.g. Unčukatli) (34). Curiously, the Turči and Shuni *govory* palatalize /k/, etc. to [č] only preceding [a] and not [i] (35).

(34)	lit.	Unč.	Kuma	Turči	gloss
	<i>huluu</i>	<i>hulu</i>	<i>hlu</i>		'lentil'
	<i>xamaa</i>	<i>xama</i>	<i>xma</i>	<i>xamaa</i>	'foam'
	<i>qu^hruu</i>	<i>qu^hru</i>	<i>qu^h</i>		'pea'
	<i>ššamaa</i>	<i>ššama</i>	<i>ššma</i>		'gravel, detritus'
(35)				<i>ččaši</i>	'hunger'
				<i>k'ič'</i>	'button'

The Shadni dialect is basically a čš/- dialect. All Shadni Lak speakers are bilingual in Dargwa, and many Dargwa loans are attested.

Khaidakov (1966) includes one final group of dialects, namely the Wachi-Kuli group. While variation and feature-mixing are typical of most (all?) Lak dialects on the level of individual *govory*, the variation attested between the two *auls* that constitute this group are sufficient to perhaps place these in separate groups, each reflecting features of different groups of Lak dialects. For example, degemination is a synchronically active process in the Wachi dialect, but not in Kuli, which exhibits the Ashti-Wixli/Arakul tendency to allow emphatics in coda position (36). While Wachi exhibits the morphophonemic [č] ~ [k]/[š] ~ [x] alternations attested in Kumux (discussed in § 48.1.2), Kuli shows only [k]/[x] in all these forms (37). Also, labialization is lacking in Wachi, which thus is like many *govory* of the Wicxi dialect,

while labialization is quite common in Kuli (38), even present in a number of words that lack it in the literary standard.

- | | | | |
|------|-------------------------|-----------------------------|------------------------|
| (36) | Wachi | Kuli | gloss |
| | <i>qqatlul</i> | <i>qqattlul</i> | ‘house.GEN’ |
| | <i>laxlan</i> | <i>laxxlan</i> | ‘to have studied’ |
| (37) | <i>lič’in (lik’lan)</i> | <i>lik’in</i> | ‘to fall asleep’ |
| | <i>lišin (lixlan)</i> | <i>lixin</i> | ‘to wipe’ |
| (38) | <i>marč</i> | <i>marč^w</i> | ‘wind’ |
| | <i>nuwč’a</i> | <i>nuwč^wa</i> | ‘flint’ |
| | <i>lešša’in</i> | <i>lešš^wa’in</i> | ‘to extinguish/go out’ |

Features shared by these dialects include the preservation of labialized [k-] sounds while non-labialized [k-] sounds shift to palatals in the relevant environment.¹⁹ Also literary [c] etc. corresponds to [č] etc. in Wachi and Kuli (39).

- | | | | |
|------|--------------------------|------------------------------------|------------------|
| (39) | lit. | (Wachi-)Kuli | gloss |
| | <i>burk^wi</i> | <i>burk^wi</i> | ‘savory pie’ |
| | <i>k^wiwa</i> | <i>k^wi^ba</i> | ‘2’ |
| | <i>kkiž</i> | <i>ččiz</i> | ‘tops (of veg.)’ |
| | <i>k’ič’</i> | <i>č’ič’</i> | ‘button’ |
| | <i>buc^wan</i> | <i>buc^wan</i> | ‘to squint’ |
| | <i>ssu’rq</i> | <i>šš^wu’rq</i> | ‘elbow’ |

In addition, a number of unpredictable and sporadic consonantal correspondences are found throughout the various Lak dialects (40).

- | | | | |
|------|---------------------------|--------------------------|--------------|
| (40) | lit. | Arakul dialect | gloss |
| | <i>cuša</i> | <i>ccuša</i> | ‘bear’ |
| | <i>qq^wa’ku</i> | <i>qq^wadu</i> | ‘raven’ |
| | <i>k’issa</i> | <i>c’issa</i> | ‘finger’ |

48.3.2. Vowels

Vowel correspondences among the various Lak dialects frequently appear in only a limited subset of lexical items in the dialects under consideration. An apparent exception to this tendency is the monophthongization of certain vowel + glide sequences in the Kuli dialect, which are realized as long middle height vowels (±pharyngealized). Perhaps the most frequent type of sporadic vowel correspondences among the dialects is between pharyngealized

19. Apparently delabialized in the Wachi variant?

and unpharyngealized vowels, or between one pharyngealized vowel and another (41). However, other vowel correspondences are attested, sometimes the result of an assimilation process (42).

- | | | | |
|------|------------------------------|--|----------------------|
| (41) | lit. | dialect | gloss |
| | <i>xurx</i> | <i>xu^ɾrx</i> (Balx.) | 'saliva' |
| | <i>buc^wan</i> | <i>bu^ɕc^wa^ɕn</i> (Balx.) | 'to tear off' |
| | <i>q'aža</i> | <i>aža</i> (Balx.) | 'goose-foot/orach' |
| | <i>č'iwissa</i> | <i>č'i^wi^ɕssa</i> (Shad.) | 'small' |
| | <i>ssur^ɕq</i> | <i>ssulq</i> (Shad.) | 'elbow' |
| | <i>ri^ɾššunak'</i> | <i>ri^ɾššunak'</i> (Balx.) | 'sour milk' |
| | <i>gikku</i> | <i>ki^ɾkku</i> (Balx.) | 'there' |
| | <i>qi^ɕt'</i> | <i>qu^ɕt'</i> (Balx.) | 'uvula' |
| | <i>č'ɪⁿpi</i> | <i>č'uⁿpi</i> (Shad.) | 'saliva' |
| (42) | <i>č'imuč'ali</i> | <i>č'imič'ali</i> (Arak.) | 'butterfly' |
| | <i>dulun</i> | <i>dulin</i> (Shad.) | 'to give (class IV)' |

A kind of umlaut or regressive spread of [-back] affecting [a] before [i], i.e. [a] > [e] / __C₀i, is characteristic of certain Lak dialects, e.g. Shadni (43) or the Katrux *govor* of the Arakul dialect (with examples such as *ššar* 'village' but GEN.SG *ššenni*). This spreads leftwards to all relevant vowels.

- | | | | |
|------|----------------------------|----------------------------|--------------|
| (43) | lit. | Shadni | gloss |
| | <i>kk^warčči</i> | <i>kk^werčči</i> | 'tooth' |
| | <i>lažin</i> | <i>ležin</i> | 'face' |
| | <i>dahani</i> | <i>deheni</i> | 'mirror' |

In addition, there are a few Kumux [a] Shadni [e] correspondences that are not harmonically (assimilatorily) motivated, e.g. *nah* : *neh* 'fat'.

The problem of the status of vowel length in the Lak dialects is a vexing one. Sources are highly contradictory and inconsistent;²⁰ for example, Arakul is described as lacking long vowels by Khaidakov (1966) but as possessing them by A. Abdullaev (1981). The mixed status of long vowels in Lak is reflected in their appearance in Russian loans alternatively in stressed and unstressed syllables, e.g. (from Murkelinski 1984) *zamāskka* but *diriktūr*.

According to A. Abdullaev (1981) long vowels are phonologically active in at least the Kuli dialect of Lak. Vowel length distinguishes a number of lexical items, occasionally reflecting some morphological process (44).

20. Due to the rather jumbled array of dialectal data in the Lak linguistic literature, this problem is echoed again and again throughout this and other Western studies of this Dagestani language.

- | | | | | |
|------|--------------|--------|--------------|---------------|
| (44) | <i>ūssa</i> | ‘born’ | <i>ussa</i> | ‘present one’ |
| | <i>āʿral</i> | ‘army’ | <i>aʿral</i> | ‘military’ |

Kuli possesses a series of long vowels of mid-height, namely *ō*, *ōʿ*, *ē*, *ēʿ*, which correspond to other dialects [aw ~ ow], [aʷ ~ oʷ], [ay ~ oy], and [aʷy ~ oʷy], respectively (45). These are not distinct from [e], [o], etc. as these latter are lacking in the Kuli dialect of Lak.

- | | | | | |
|------|-----------------|--------------|-------------|--------------|
| (45) | lit. Lak | Kumux | Kuli | gloss |
| | <i>ssaw</i> | [ssow] | <i>ssō</i> | ‘sky’ |
| | <i>ššaʷw</i> | [ššoʷw] | <i>ššōʿ</i> | ‘land.LOC’ |
| | <i>may</i> | [mey] | <i>mē</i> | ‘nose’ |
| | <i>taʷy</i> | [teʷy] | <i>tēʿ</i> | ‘foal’ |

Other sources for long vowels in (Kuli) Lak include the monophthongization of /iw/ and /uy/ to [ii] (46). A similar phenomenon is seen in the formation of certain plurals and verbal aspect forms and various other derivationally related sets of lexical items (47); in the perfect forms, infix -w- has coalesced to form a long vowel, thus creating a grammaticalized length opposition.

- | | | | |
|------|-----------------|-------------|-----------------|
| (46) | lit. Lak | Kuli | gloss |
| | <i>makiw</i> | <i>maḱī</i> | ‘roofs’ |
| | <i>lamuy</i> | <i>lamī</i> | ‘on the bridge’ |
-
- | | | | | |
|------|----------------|-------------|----------------|-----------------|
| (47) | <i>litʷun</i> | ‘to kill’ | <i>lītʷun</i> | ‘having killed’ |
| | <i>issun</i> | ‘redeem’ | <i>īssun</i> | ‘redeemed’ |
| | <i>buhu</i> | ‘fisherman’ | <i>būhu</i> | ‘caught’ |
| | <i>lōgʷuna</i> | ‘left’ | <i>lōgʷunā</i> | ‘leave.GER’ |
| | <i>χχʷanu</i> | ‘peg; rack’ | <i>χχʷanū</i> | ‘PL’ |
| | <i>buʷqʷru</i> | ‘drowning’ | <i>būʷqʷru</i> | ‘they drowned’ |

A. Abdullaev (1981) makes a contrast not only between short and long vowels in the Kuli dialect, but also between long vowel and geminate vowel (48). Between the two identical vowels in the geminate is a [ʔ].

- | | | | | |
|------|---------------------------|---------------------|-----------------|----------------------|
| (48) | <i>taar</i> ²¹ | ‘that meadow’ | <i>tār</i> | ‘musical instrument’ |
| | <i>ttiizunni</i> | ‘suddenly stood up’ | <i>ttīzunni</i> | ‘they got milk’ |

In sum, vowel length and stress are mutually dependent in some Lak dialects; in others, e.g. Kuli, vowel length is a fairly vital part of the (morpho)phonology, due largely to monophthongization of Pre-Lak diphthongs. However, due to the general inconsistency in identifying forms, as well as the

21. Presumably with cliticized demonstrative/deictic.

sometimes contradictory claims in the literature (e.g. the presence or absence of vowel length attributed to the Arakul dialect by various authors), a more accurate assessment of the status of vowel length in the various dialects requires further research. The same can also be said about the status of phonemic labialized obstruents and stress assignment in Lak.

References

- Abdokov, A. I. 1983. *O zvukovykh i slovar'nykh sootvetsviakh sever-okavkazskikh iazykov*. Nal'chik: "Elbrus."
- Abdullaev, A. A. 1981. "Fonologicheskaja priroda dolgoty-kratkosti glasnykh lakskogo iazyka (na materiale govora aula Kuli)." In *Foneticheskaja sistema dagestanskikh iazykov*, pp. 72–85. Maxačkala: Dag. Fil. AN SSSR.
- Abdullaev, I. X. 1981. "O metateze i kompleksakh soglasnykh v lakskom iazyke." In *Foneticheskaja sistema dagestanskikh iazykov*, pp. 38–53. Maxačkala: Dag. Fil. AN SSSR.
- . 1987. "Ob ablaute v sisteme lakskogo imennogo skloneniia." In *Pa-dezhnyi sostav i sistemy skloneniia v kavkazskikh iazykakh*. Maxačkala: Dag. Fil. AN SSSR.
- . 1992. "Dialektnye variatsii mestoimeni osnovy wa "etot" v lakskom iazyke." In *Dialektologicheskoe izučenie dagestanskikh iazykov*. Maxačkala: Dag. nauch. tsentr. inst. iaz., lit., ist. im. G. Tsadasy.
- Bokarev, E. A. 1961. *Vvedenie v sravnitel'no-istoricheskoe izučenie dagestanskikh iazykov: Material k kursu*. Maxačkala: Dag. Gos. Univ. im V. I. Lenina.
- . 1981. *Sravnitel'no-istoricheskaja fonetika vostochnokavkazskikh iazykov*. Moscow: Nauka.
- Bouda, Karl. 1949. *Lakkische Studien*. Heidelberg: Winter.
- Burčulaje, G. T. 1984. "Ob ablaute glasnykh v imenakh sushchestvitel'nykh lakskogo iazyka." *Iberiul ɣavɣasiuri enatmecnierebis čeličdeuli (IKEC)* 11: 128–36.
- . 1985a. "Čvenebit nacvalsaxelta pujeebi laɣurši." *Iberiul ɣavɣasiuri enatmecniereba* 23: 301–7.
- . 1985b. "K istorii konechnykh glasnykh imen sushchestvitel'nykh v lakskom iazyke." *IKEC* 12: 286–308.

- Burzhunov, G. G. 1984. "Foneticheskoe osvoenie russkikh zaimstvovaniia dagestanskimi iazykami i ego znachenie v razvitiu dviiazychiia." In *Russkii iazyk i ego vliianie na razvitie dagestanskikh iazykov*, pp. 55–67. Maxačkala: Dag. Fil. AN SSSR.
- Catford, J. C. 1994. "Vowel Systems of Caucasian Languages." In *Non-Slavic Languages of the USSR: Papers from the Fourth Conference*, ed. Howard I. Aronson, pp. 53–54. Columbus, Ohio: Slavica.
- Cercvadze, I. I. 1980. "E da o xmovanta saķitxisatvis laķurši." *Iberiul-ķavķasiuri enatmecnierebis ķeliķdeuli* 7: 228–31.
- Dirr, Adolf. 1928. *Einführung in das Studium der kaukasischen Sprachen*, pp. 238–53. Leipzig: Verlag der Asia Minor.
- Erckert, R. von. 1895. *Die Sprachen des kaukasischen Stammes*. Repr. Wiesbaden: Sändig, 1970.
- Gaprindashvili, Sh. 1954. "Bgeratšesaťqvisobebisatvis darguulsa da laķurši." *Iberiul-ķavķasiuri Enatmecniereba* 6: 281–327.
- Gigineishvili, B. K. 1977. *Sravnitel'naia fonetika dagestanskikh iazykov*. Tbilisi: Tbilisi Gosudorstvennyi Universitet.
- Hapelsmath, Martin. 1993. *A Grammar of Lezgian*. Berlin: Mouton / de Gruyter.
- Hewitt, B. G. 1981. "Caucasian Languages." In *The Languages of the Soviet Union*, by Bernard Comrie, pp. 196–237. Cambridge: Cambridge University Press.
- Khaidakov, S. M. 1962. *Lakku mazral wa u'rus mazral slovar'*. Moscow: Gos. izd. inostrannykh i natsional'nykh slovarei.
- . 1966. *Ocherki po lakskoi dialektologii*. Moscow: Nauka.
- . 1981. "Mesto ablautnogo cheredovaniia v slovoizmenenii dagestanskikh iazykov." In *Foneticheskaia sistema dagestanskikh iazykov*, pp. 3–13. Maxačkala: Dag. Fil. AN SSSR.
- Kibrik, A. E., and S. V. Kodzasov 1990. *Sopostavitel'noe izuchenie dagestanskikh iazykov: imia i fonetika*. Moscow: izd. Moskovskii Gosudorstvennyi Universitet.
- Murkelinski, G. B. 1949. "Kratkie svedeniia o vixlinskom dialekte lakskogo iazyka." In *Iazyki Severnego Kavkaza i Dagestana*, pp. 86–102. Moscow: AN SSSR.
- . 1967. "Lakskii iazyk." In *Iazyki narodov SSSR*, vol. 4: *Iberiisko-kavkazskie iazyki*, pp. 488–507. Moscow: Nauka.
- . 1971. *Grammatika lakskogo iazyka*. Makhachkala: Daguchpedgiz.

- . 1984. "O vliianii russkogo iazyka na razvitie laskkogo literaturnogo iazyka." In *Russkii iazyk i ego vliianiie na razvitie dagestanskikh iazykov*, pp. 115–28. Maxačkala: Dag. Fil. AN SSSR.
- Talibov, B. B. 1977. "O processe oglusheniia zvonkikh soglasnykh v lezgin-skikh iazykakh." *IKEC* 4: 248–57.
- Trubetzkoy, N. S. 1931. "Die Konsonantensysteme der ostkaukasischen Sprachen." *Caucasica* 8: 1–52.
- Uslar, P. K. 1890. *Lakskii iazyk* (Ètnografia Kavkaza Iazykoznanie 4). Tbilisi: Upravlenie Kavkazskogo Učebnova Okruga.
- Zhirkov, L. I. 1955. *Lakskii iazyk: fonetika i morfologiya*. Moscow: Izdat. AN SSSR.

This page intentionally left blank.

8 Unaffiliated Languages (Language Isolates)

This page intentionally left blank.

Sumerian Phonology

John Hayes

University of California, Berkeley

49.1. External history

Sumerian was the language spoken by the Sumerians, who inhabited the southern part of ancient Mesopotamia (present-day Iraq). Sumerian has the distinction of being the first attested language in history: its earliest texts date to about 3100 B.C.E. There is no agreed-upon periodization of the language, and the periodizations in use combine both linguistic criteria and non-linguistic criteria—such as political and other historical events—to outline its external history. One such periodization is:

Archaic Sumerian	3100–2600 B.C.E.
Classical Sumerian	2600–2300 B.C.E.
Neo-Sumerian	2300–2000 B.C.E.
Post-Sumerian	2000 B.C.E.–100 C.E.

It is not known when Sumerian died, that is, when it ceased to be spoken as a first language. This is a thorny issue much discussed among Sumerologists. Many believe that for all intents and purposes it was dead in 2000 B.C.E. Others see a later date, perhaps around 1600 B.C.E.; others opt for an even earlier date. However, Sumerian culture was inherited by the Akkadians, a Semitic-speaking people who moved into Mesopotamia sometime after the Sumerians did. The Akkadians studied Sumerian language and literature in their school system, and composed texts in Sumerian a thousand years after its death.

The description which will be given here reflects the Classical Sumerian and Neo-Sumerian periods. The language of the Post-Sumerian period was heavily influenced by Akkadian, particularly in its syntax. For example, the ergative nature of Sumerian was not always understood by the Akkadians, who spoke a nominative–accusative language, and so the Akkadian-speaking scribes occasionally treated Sumerian as if it were a nominative–accusative language. Late texts exhibit a substantial number of “errors.” Sometimes, however, it is not known if unexpected features occurring in Neo- and Post-Sumerian texts are errors of the scribes, or whether they are

genuine features of Sumerian which were preserved in the scribal academies, but which are not yet understood.

49.2. Genetic relationships

The consensus of opinion is that Sumerian is a language isolate. Numerous attempts by both amateurs and professionals have been made to connect Sumerian with other languages; these have ranged from languages spoken in Polynesia, to Hungarian, and to Basque. It is most likely that the relatives of Sumerian died out without having been preserved in writing.

The original homeland of the Sumerians is not known; according to their own mythology, they were not indigenous to Mesopotamia. It is sometimes thought that they came from somewhere on the Indian subcontinent, and are perhaps connected with the ancient civilizations of Mohenjo-Daro and Harappa. However, there is as yet no archaeological evidence in India to support this. The attempts which have been made to relate Sumerian to the Dravidian languages have not yet produced satisfying results.

Very recently, scholars interested in long-range classification and reconstruction have posited a relationship within the Nostratic or Eurasiatic families (both of which are variously defined), but in the present stage of our knowledge such attempts are ill conceived.

49.3. Typology

In older schemes of typological classification, Sumerian is typically listed among the “agglutinative” languages, primarily because it has a series of postpositions to indicate case relationships.

Word order is basically SOV, and adjectives follow the noun. Two types of genitive construction occur; one is the construction typical of SOV languages, the other (more common) is not.

The verb appears to encode for aspect and not tense, although this point has been argued; in royal inscriptions, for example, the system seems more tense-like than aspect-like.

Sumerian operates on a split ergative basis: the perfect aspect is ergative, and the imperfect aspect is nominative–accusative. A series of verbal prefixes cross-references the various persons of the verb. While the broad picture of how these prefixes are correlated with ergativity is understood, many details of ergativity in Sumerian are still unsure. Syntactic ergativity has hardly been studied.

49.4. External influences

The Sumerians and the Akkadians lived together peacefully for centuries; the Sumerians themselves were eventually assimilated into the Akkadians and other Semitic-speaking peoples. This led to a heavy influence of Sumerian upon Akkadian, in all aspects of the language. For example, Akkadian has shifted from a presumed VSO syntax (to argue from comparative evidence) to an SOV order, very uncommon in Semitic; Ethio-Semitic has undergone a similar change in syntax, under the influence of the Cushitic languages. Akkadian has also borrowed a very large number of words from Sumerian.

Virtually nothing is known about the language(s) spoken in Mesopotamia prior to the arrival of the Sumerians. A fair number of place names and common nouns (*apin* ‘plow’; *nagar* ‘carpenter’) are probably substrate words.

Akkadian influence upon Sumerian shows up early; the Akkadian conjunction /u/ ‘and’ appears in Sumerian as early as 2600 B.C.E. Akkadian influence was quite strong in late periods, particularly when Sumerian had died out as a spoken language.

49.5. Dialects

Only a limited amount of what may be geographical variation is observable in Sumerian texts. Sumerian was spoken in a relatively small area, and the logographic nature of the writing system masks possible geographical differences. More geographical variation occurs in relatively late texts, but some features commonly ascribed to geographical variation within Sumerian may reflect differences in the practices of scribal schools, differences rooted in Akkadian, not in Sumerian.

Sumerian texts occasionally refer to such “languages” as “the language of sailors” or “the language of shepherds,” but these terms probably refer to the specialized vocabulary of certain professions; nothing is known of these argots except their names.

In addition to “standard” Sumerian (which is commonly referred to as “Main Dialect”), there are a number of texts (and words embedded inside Main Dialect texts) which are written in a sociolect called *eme-sal*. *Eme* is the Sumerian word for ‘tongue, language’; it is not sure what *sal* means in this context. *Emesal* appears to be a “women’s language.” It is used primarily by goddesses when speaking to other goddesses; it is also occasionally used by certain kinds of priests, probably eunuchs, for specific types of incantations. Curiously, when the actual speech of women (as opposed to that

of goddesses) is preserved, as in certain kinds of legal texts, Main Dialect is used, not Emesal.

Other interpretations have been made—for example, that it represents an “ingratiating style,” as opposed to a “women’s language.” There has been less discussion about the “origins” of Emesal, although it has recently been speculated that it arose as a geographical dialect in the south of Sumer which became closely tied to the temple and cult of the goddess Inanna.

It is not easy to give a picture of Emesal, because the largely logographic nature of the Sumerian script hides Emesal pronunciation. Moreover, scribes often wrote Main Dialect forms even when these forms were meant to be pronounced in Emesal. This means that often the Emesal “reading” of a word is unknown. The total corpus of Emesal material is relatively small, and attested rather late, further serving to restrict knowledge of Emesal.

Emesal is more conservative than Main Dialect in its morphology and its lexicon, yet at the same time shows some innovative tendencies in its phonology. Such a distribution of conservative–innovative features is said to be typical of women’s languages.

49.6. Writing system

Any reconstruction of Sumerian phonology (and of Sumerian grammar in general) immediately runs into the problem of the nature of the script which was used to write Sumerian. Most Sumerologists believe that the Sumerian script was invented by the Sumerians; a minority view is that the script was created by some other people, whose identity is now lost to us. In any case, in origin the Sumerian script was largely a mnemonic device. It did not attempt to fully reproduce actual speech. The earliest uses of writing were for accounting and administrative texts; these texts largely served to jog the memory of the scribes, who were naturally familiar with the contexts of the texts, and saw no need to write down every phoneme or even every morpheme. In the earliest Sumerian, basically only lexical morphemes were written, and only occasionally grammatical morphemes. For example, the sign representing the lexeme ‘to build’ could be used to represent essentially any inflected form of the verb, that is, the categories of person, number, and gender were not graphically marked. As time passed, more and more grammatical information began to be written down. However, the script **never** was a full representation of speech.

The Sumerian writing system is a mixed logographic–syllabic system. In general, logographic writings are used to represent lexical morphemes, and

syllabic spellings to represent grammatical morphemes (and occasional loanwords). The logographic nature means that it is very difficult to determine precise phonetic information, or to investigate diachronic change.

Many signs are polyvalent, that is, they have more than one value. Some signs, for example, have more than one logographic value; others have more than one syllabic value; some signs have both several logographic and several syllabic values. The same sign, for example, can be read both as /bi/ or as /be/. Such polyvalency hinders our knowledge both of the morphology and of the phonology.

To some extent the Sumerian writing system is morphographemic, in the sense that morphemes were occasionally written in their fullest, underlying, form, even if in certain phonological contexts phonetic change or reduction took place. Thus, the postposition marking the dative case was occasionally written *ra*, even if pronounced /r/.

49.7. Transliteration and transcription

Sumerian is typically cited in two different ways. **Transliteration** is a sign-by-sign reflection of the cuneiform; its purpose is to enable scholars to determine exactly which signs were written on the tablet (exclusive of palaeographic niceties). **Transcription** is an attempt to indicate the presumed pronunciation of a complete word, and usually includes morphemes assumed to be present, but not actually written down in the text. For example, a transliteration such as *mu-un-na-dù* shows that four signs occur on the tablet, and states explicitly which four signs. These same four signs might be transcribed as *munandu* ‘he built’. This transcription shows that the gemination of the first *u* (*mu-un*) is graphic only; that the gemination of the first *n* (*un-na*) is also graphic only; and that there is a pronominal element /n/, not shown in the writing, immediately before the verbal root /du/.

Transliteration is thus sign-by-sign, while transcription is typically word-by-word. Transliterations will essentially be the same for all scholars. Since transcription, however, reflects the presumed pronunciation of the entire word, transcriptions may differ from scholar to scholar, depending on individual interpretations of the writing system and of the morphology.

Because of the large number of homophonic signs in Sumerian (there are a number of signs, for example, all pronounced as /bi/), scholars have devised an arbitrary system to enable other scholars to determine which of these signs were actually present in a text. This is done by a set of diacritics (or “indices”). The most common (or important) value of a sign is

unmarked. The second most common is marked by an acute accent: *bí*. The next most common is marked by a grave accent: *bì*. Others are marked by subscripts: *bi₄*, *bi₅*, etc. These indices convey no phonetic information. They do not mark different vowels, or stress, or tones.

Because so much of our difficulty in understanding Sumerian comes from trying to determine the reality behind the writing system, Sumerian will normally be cited here in transliteration.

49.8. Sources of knowledge

If, then, the Sumerian language has completely died out, leaving no known relatives living or dead, and if the writing system is often opaque, in that key features of the morphology are simply not written down, how can Sumerian be understood? Part of the answer lies in the fact that the Akkadians, the inheritors of Sumerian civilization, adopted the Sumerian writing system for their own language, Akkadian, a well-understood Semitic language with numerous cognate languages living and dead. This means that we can project Akkadian syllabic readings of individual signs back onto Sumerian, and thus obtain the reading of Sumerian syllabic signs. Akkadian also borrowed many words from Sumerian, and so Sumerian words written logographically in Sumerian are sometimes spelled out syllabically in Akkadian. Unfortunately, however, our knowledge of the phonetic structure of Akkadian is less sure than commonly assumed.

A further source of knowledge is the fact that the Akkadians studied Sumerian in their scribal academies. They compiled numerous vocabulary lists, which often spell out the approximate pronunciation and the meaning of Sumerian words; variant Akkadian spellings of Sumerian loanwords are one entry into the pronunciation of Sumerian. The Akkadians studied Sumerian grammar, composing lists of grammatical morphemes with their equivalents in Akkadian. They copied and recopied Sumerian literature, glossed Sumerian texts, and occasionally wrote interlinear translations into Akkadian. If the Akkadians had not borrowed the Sumerian script and had not actively pursued its study, our knowledge of Sumerian would be much more limited.

Finally, a few Sumerian words (especially proper names) have been borrowed into other languages and other writing systems. For example, the Sumerian word for 'son' *dumu* appears in a very late Greek spelling as δομ. It is very difficult to utilize such information, which is generally scattered and

often late. Often the exact phonetic structure of the language in which the transliterated Sumerian word occurs is not precisely known, and this lack of knowledge can lead to circular reasoning.

49.9. State of knowledge

In the most general sense, Sumerian is well understood, in that relatively straightforward texts can easily be interpreted and translated. Difficult texts, such as poetry and literature in general, can be much more challenging, and translations produced by competent scholars can occasionally differ widely. The opaqueness of the writing system in particular makes it difficult for us to understand all the features of the morphology. This is particularly true for the pronominal system and the various modal markers. It is often possible to explain in broad terms which is happening in a text, but impossible to explain all details of the grammar.

At times, it is difficult to decide whether a problem in understanding a particular Sumerian form is to be solved at the graphic level, at the phonological level, or at the morphological level. For example, the suffixed demonstrative *-bi* ‘this’ when followed by *-e*, the marker of the ergative case, is written *-bi*. How is this to be understood? Should the sign be “read” as *-bé*, implying a contraction /bie/ > /be/ or /bē/, or should it be read as *-bi*, implying either a contraction /bie/ > /bi/ or /bī/, or is the ergative marker to be understood as *-Ø* after a vowel?

It is precisely in the field of phonetics and phonology where our knowledge of Sumerian is weakest, and some basic facts are unclear. This means that it is not possible to give an integrated description of Sumerian phonology, and the discussion which follows will perforce be somewhat atomistic; major areas of disagreement will be noted.

49.10. Phonemic inventory

As may be inferred from the above discussion, we cannot yet completely recover the Sumerian phonological system, or determine the precise phonetic nature of its posited phonemes. On the one hand, the interest of the Akkadians in studying Sumerian enables us to obtain a general picture, yet at the same time our knowledge of Sumerian is filtered through our knowledge of Akkadian. For example, it is quite possible that Sumerian had an /o/-quality vowel. Akkadian did not, however, and since our understanding

of the writing system is largely based on our knowledge of Akkadian, we do not clearly see an /o/-quality vowel, and the evidence for its existence is only indirect.

As stated above, much of our knowledge of Sumerian results from our knowledge of the syllabic readings of (Akkadian) cuneiform signs. Unfortunately, these syllabic spellings are almost uniformly late, from the Post-Sumerian period, and in fact these syllabic signs largely developed in the Akkadian scribal system. This means we are projecting Akkadian values backward. Moreover, there undoubtedly was an oral component in the Akkadian school system, which we can never retrieve, which conceivably would have explained to the Akkadian students and scribes differences between Sumerian and Akkadian.

49.10.1. *Consonants*

Table 49-1 lists the consonants whose existence in Sumerian is uncontroversial.

Table 49-1. Sumerian Consonants

b	p	m
d	t	n
g	k	ḡ
z	s	š
	ḥ	
l	r	

ḡ is the standard Sumerological transcription of the velar nasal [ŋ]; ḥ is the uvular fricative [x]; š the palato-alveolar fricative [ʃ].

The writing system has no unambiguous way to indicate the consonants /y/ and /w/, but a word such as *a-a* ‘father’ presumably represents /aya/. The evidence for /w/ is less clear.

Regarding the stops, the nature of the opposition presented above in terms of voice (*b ~ p, d ~ t, g ~ k*) has been much discussed. While the difference in Akkadian was one of voice (to argue from comparative evidence), it is probable that in Sumerian the difference was one of aspiration. The series traditionally transliterated as the voiceless stops *p t k* should be interpreted as the voiceless aspirates /p^h t^h k^h/, and the series traditionally transliterated as the voiced stops *b d g* should be interpreted as the voiceless non-aspirates /p t k/. This produces a language with two sets of stops: voiceless aspirates

and voiceless non-aspirates; such systems are not uncommon in the languages of the world.

The evidence for this opposition partially derives from the way Sumerian words were borrowed into Akkadian. At least in the early periods of Sumerian, in general Sumerian words with initial (and sometimes medial or final) “voiced” stops appear in loanwords into Akkadian as voiceless stops: Sumerian /barag/ ‘throne dais’ > /parakku/; *gala* ‘kind of priest’ > *kalû*. Similarly, early Akkadian loanwords into Sumerian show the opposite phenomenon: Akkadian *tamḥāru* ‘battle’ appears as Sumerian *dam-ḥa-ra*. This phenomenon is regular in initial position, less so for consonants in medial and final position.

This interpretation of the opposition between the two series of stops is the most common understanding today, although transliterations only rarely reflect this interpretation. This means that the actual phonetic shape of Sumerian may be quite different than standard transliterations show; for example, the standard transliteration *gibil* ‘new’ may actually represent /kipil/.

The values for the sibilants are derived from the traditional interpretation of the pronunciation of the Akkadian sibilants, but since the question of the sibilants in Semitic and in Old Akkadian is still unresolved, these conventional transliterations (*z*, *s*, *š*) should not be taken as any accurate phonetic rendering. Claude Boisson, for example, interprets *z s š* as /t^s s θ/.

There is little evidence to indicate the type of /r/ Sumerian may have had. The Akkadian word *šuršu* ‘root’ appears syllabically in Sumerian as *šu-ḥu-uš*, and the Sumerian word for ‘red’, *ḥuš-a*, appears in Akkadian both as *ḥuššû* and *ruššû*. This may indicate that the Sumerian /r/ was uvular, and not a flap.

The only phoneme of the above phonemic inventory which does not occur in Akkadian is /ḡ/. Its existence is established by spellings which oscillate between *m*, *n*, *g*, and *ng*. Some have argued for a more complex realization, such as a labialized velar nasal /ḡ^w/, or a nasalized labio-velar. (The symbol *ḡ* is used to avoid specifying the exact phonetic realization.)

Another phoneme whose existence is widely accepted is usually transcribed /dr/ (sometimes as /d^r/). Its precise phonetic value has seldom been discussed, though it is widely regarded as some kind of flap. Its existence is based on variant spellings which show it to be a sound distinct from /d/, /t/, and /r/. The verb meaning basically ‘to cut’, for example, sometimes appears in syllabic spellings with *d*, sometimes with *t*, sometimes with *r*. The existence of this phoneme has only been shown for a small number of lexical morphemes.

All Sumerologists accept the existence of /ḡ/, and most the existence of /dr/ (even if the phonetic nature of these two consonants is not precisely specified). Other consonants have also been postulated, but the evidence is more tenuous. The evidence is usually predicated on odd spellings and on the behavior of Sumerian loanwords into Akkadian. It is possible that some of the theoretically reconstructed consonants are allophones, and not phonemes, in Sumerian. Possible additions to the consonantal inventory include the following:

/h/ This is accepted by many, but the only evidence cited is the fact that the Sumerian word *é-gal* 'house-big' > 'temple' appears in various Semitic languages with an initial /h/: Ugaritic *hkl*, Hebrew and Aramaic *hekal*, Arabic *haykal*. But this would imply that the Ugaritic, Hebrew, and Aramaic forms were borrowed directly from Sumerian, yet as far as we can tell there was no contact between speakers of Sumerian and speakers of any of these languages; the Sumerian loanwords occurring in the various Semitic languages were all transmitted through Akkadian, which does not have an /h/. (If the opposition of stops discussed above were one of aspiration ~ non-aspiration, one might expect to find /h/ in the phonemic inventory, but not necessarily so.)

/l/ Certain words ending in -/l/ are always written with the *la* sign when followed by an /a/-vowel in certain grammatical constructions, while other words ending in -/l/ are written with the *lā*-sign when followed by an /a/-vowel in certain grammatical constructions. This distribution may mean that there were two types of /l/, but it is not possible to establish the phonetic difference.

/r/ Similar spelling conventions in final position may imply the existence of two types of /r/. Similarly, in initial position *d* occasionally alternates with *r*; implying two /r/s. Several theoretical phonetic varieties are possible.

/ɣ/ The modal prefix for the cohortative, usually written -*ga*, is sometimes written as *ḡa*. This may indicate that at least some of the /g/s in Sumerian were closer to /ɣ/ or some similar phoneme. There is no other evidence for the existence of such a phoneme.

/g^w/ The Sumerian word *agrig* 'a kind of steward' appears in Akkadian as *abarakku*, hinting that the initial /g/ was a consonant more complex than a simple voiced velar stop, perhaps /g^w/ or a unit phoneme /gb/. (In Eblaite, another Semitic language attested very early, the word is spelled both *à-ga-ra-gú-um* and *à-ba-ru₁₂-gú*). Other cases seem to occur. As pointed out by Boisson, however, the presence in the Sumerian consonantal inventory of /g^w/ or

of /gb/ would almost certainly entail the presence of /k^w/ or /kp/, and there is no clear evidence for this.

Transliterations of Sumerian texts differ in the way they indicate the phonemes which do not occur in Akkadian but which have been postulated for Sumerian. Typically *ḡ* is transliterated as such, but sometimes not; the first person possessive suffix /ḡu/, for example, is sometimes transliterated as *-ḡu*₁₀ and sometimes as *-mu* (the most common value of this sign in Akkadian). The other possible phonemes discussed above are not usually transliterated, but rather the values known from Akkadian are used instead; an edition of a Sumerian literary text, for example, will typically show only phonemes known in Akkadian. Transliterations of these disputed phonemes usually occur in discussions by scholars particularly interested in Sumerian phonology or morphology.

There is no clear evidence that consonantal length was a phonemic feature in Sumerian; there are no obvious minimal pairs. Long consonants can arise from suffixation or occasionally from assimilation (**bar-bar* > *babbar*). They also occur in certain lexical items, but the standard transliterations of lexical items with long consonants may not always be correct. Transliterations of syllabically written signs, such as prefixes on verbs, frequently show graphic long consonants, but this is because Sumerian follows a graphic convention whereby sequences of the type /C₁V₁-C₂V₂/ are represented by a sequence of three signs: C₁V₁-V₁C₂-C₂V₂ (*mu-un-na* = /muna-/). Similarly, Sumerian follows a convention whereby if a suffix beginning with a vowel is added to a word ending in a consonant, the consonant is graphically reduplicated; a transliteration such as *kalag-ga* is to be understood as /kalaga/, representing a root *kalag* ‘to be strong’ plus a nominalizer /a/.

49.10.2. Vowels

Sumerian had at least the vowels shown in Table 49-2.

Table 49-2. Sumerian Vowels

i	u
e	
a	

Needless to say, the precise phonetic character of these vowels is not known. As mentioned above, Sumerian may well have had an /o/-quality vowel. However, Akkadian had only the same four vowels shown in the

chart above, so it is very difficult to demonstrate the existence of other vowels in Sumerian. The evidence for /o/ is based on spelling variations and late transcriptions into other languages. Because of the limited nature of the evidence, its existence has been shown for only a few words; it is virtually impossible to tell if a word contains /o/ or /u/. The presence of this vowel would yield a five-vowel system as in Table 49-3.

Table 49-3. Possible Sumerian Five-Vowel System

i		u
	e	o
	a	

Other vocalic systems have been proposed, including, for example, a series of umlauted vowels: /ä, ï, ö, ü/. The most recent proposal, that of Bobrova and Militarëv, suggests an eight-vowel system, which they chart as in Table 49-4.

Table 49-4. Possible Sumerian Eight-Vowel System

		o	
a			u
ä			ü
	i	e	
	ə		

The presence of some of these vowels is inferred from variant spellings. If a word is spelled, for example, with the vowel /u/ one time, and with the vowel /i/ another time, this might mean that the vowel was an intermediate vowel of some kind, perhaps /ü/, a vowel which does not exist in Akkadian, so Akkadian speakers sometimes interpreted it as /u/, sometimes as /i/ (this alternation of vowels is the approach used by Bobrova and Militarëv to generate the eight-vowel system in Table 49-4). The problem with this method is that such spellings typically are late, coming from a period when Sumerian was no longer a spoken language, and in fact it is very difficult to find such alternations within one specific time and place. It also leads to a confusion between the description of Proto-Sumerian and that of historical Sumerian.

It is not known if both short and long vowels existed, at the phonemic level or otherwise; the writing system cannot unequivocally show vocalic

length. It is possible that long vowels existed as a secondary development, arising from the contraction of diphthongs or other vocalic contraction.

Because the script does not clearly indicate /y/ or /w/, it is difficult to prove the existence of diphthongs; /ay/ most probably existed, but the evidence is less strong for /aw/.

It has several times been suggested that Sumerian possessed nasalized vowels, but the number and nature of these is subject to disagreement. Some have suggested only a nasalized /ĩ/; others have suggested both /ĩ/ and /ẽ/; others /ĩ/, /ẽ/, and /ä/. Not all scholars accept their existence, and those who do posit their existence only in certain specific morphological forms. The case most often cited occurs in the prefix-chain of the Sumerian verb, where an assumed underlying /i-ba/ is written *im-ma*. It is not, however, positive that this is the correct underlying form, and other explanations are possible. The existence of nasalized vowels is not accepted here.

Transliterations of Sumerian typically use only the four short vowels *a e i u*.

49.10.3. *Typology*

Because of the fact that neither the complete consonantal or vocalic inventories are known, it is difficult to make unequivocal statements about the typological significance of the Sumerian phoneme inventory, and in fact it is only recently that this subject has begun to be investigated. The most recent study is by Claude Boisson, who examines the different vocalic systems posited by scholars, the nature of the opposition between the voiced and unvoiced consonants, and some of the individual phonemes reconstructed for Sumerian.

49.11. Other features

49.11.1. *Stress*

Very little is known about stress. Writings occur with unexpected vocalic loss, and some of these instances may indicate a strong stress on the following (or preceding) syllable. The divine name *Amar-utu-(k)* ‘Bull of the Sun’ appears in Akkadian as *Marduk*, presumably indicating an original stress /amárutuk/. A fair number of such instances occur, but generally from different places and periods, so that it is not yet possible to determine the nature of stress in Sumerian, or to determine whether the placement of stress is rule-governed or lexical.

49.11.2. *Tones*

Because Sumerian seems to have a large number of homophones, it has frequently been argued that Sumerian had a tonal system of some kind; otherwise, this abundance of homophones would have rendered communication impossible. For example, the *Pennsylvania Sumerian Dictionary* lists the following entries under *ba*: A (a tool); B ‘snail’(?) ; C ‘allotment’; D ‘to allot’. The monosyllable /u/ includes words ranging from ‘ten’ to ‘plant’ to ‘this’ to ‘to ride’. This high degree of homophony is said to result from the fact that Sumerian possesses a relatively small consonantal inventory and a small vocalic inventory, coupled with a tendency towards monosyllabic lexical roots.

From a viewpoint of general linguistic theory, there seems to be no actual hard data about how much homophony a language actually permits. Also, it may be that many of what are usually considered to be homophones in Sumerian were not actually such; there may have been phonetic differences which cannot readily be seen. Some may differ in the presence of an underlying final consonant, which is dropped under certain syllabic conditions; this is discussed below.

49.11.3. *Syllable structure*

Again because of the way our knowledge of Sumerian is filtered through Akkadian, it is not possible to fully determine Sumerian syllabic structure. At least the following syllable types occur: V, CV, CVC. These are essentially the same syllable types present in Akkadian.

It is usually thought that initial and final clusters do not occur. On the assumption that certain signs of the type C_1VC_2V stand for C_1C_2V , it has been speculated that a few lexical morphemes do in fact have an initial cluster, but that this is masked by the script, and the fact that Akkadian (and Semitic in general) does not permit initial clusters. There is as yet no conclusive evidence.

49.12. Phonological changes

49.12.1. *Consonantal loss*

Certain consonants are regularly deleted in word-final position. ‘Throne dais’, for example, underlyingly /barag/, was pronounced /bara/ at the end of a sentence. However, ‘at the throne dais’, using the postpositive locative case marker /a/, was realized as /baraga/. The consonants which are subject to this

loss are referred to in Sumerological parlance as “amissable,” and the system is known as “amissability.”

Because of the general tendency to write lexical morphemes by logograms, it is not actually easy to see amissability in practice, and a few scholars have argued that the phenomenon did not exist, but rather reflects vagaries in the writing system. Assuming the existence of amissability, two questions are unresolved. The first is, which consonants are amissable? Here there is no unanimity of opinion, and different scholars have produced different lists. The second question is the scope of amissability. Does it only apply to word-final consonants, or does it also apply more generally to syllable-final position, even within the word? The evidence for the latter is much more difficult to evaluate.

It has been pointed out that the reality behind amissability is more complex than usually thought. While a language such as French has occasionally been cited as having an amissability rule of sorts, the rule in French works across the board, while the situation in Sumerian is less easy to describe; it has been stated, for example, that /b/ is never lost, /d/ is always lost, and /g/ is mostly lost.

In cases of reduplication, which plays many roles in Sumerian morphology, there is also loss of syllable-final consonants. Typically a syllable such as *dag* ‘white’ when reduplicated produces *dadag* (also ‘white’, perhaps originally emphatic, ‘very white’). The basic rule is that C_1VC_2 syllables when reduplicated $> C_1VC_1VC_2$. Other types of reduplication produce the same consonantal loss. Exceptions also occur; another word for ‘white, shining’ *babbar* comes from **bar-bar*.

A few instances exist where a Sumerian word borrowed into Akkadian loses its initial consonant. It is not known if these represent unexplained intra-Sumerian developments, or are intra-Akkadian developments. Examples include Akkadian *eṭemmu* ‘spirit of the dead’ from Sumerian *gidim* (perhaps [kitim]) and *itinnu* ‘builder’ from Sumerian *šitim*.

49.12.2. Consonantal change

The modal prefix for the negative, *nu*, is subject to a rather unusual change:

<i>nu-ba</i>	$> la-ba$
<i>nu-bí</i>	$> li-bí$

Somewhat similarly, the Sumerian word *nu-banda*, ‘steward, lieutenant, mayor’ appears in Akkadian as *laputtû*, implying that the Sumerian was actually pronounced /labanda/ or /lapanta/. A few other cases occur, most

involving a root containing /b/. This is a curious alternation, and it is difficult to understand the motivation. It is possible that the /b/ was a more complex phoneme, perhaps nasalized, and so this is a type of dissimilation.

49.12.3. *Vocalic assimilation*

At some point prior to our first written records, Sumerian underwent a fairly extensive process of assimilation of vowels in adjacent syllables within words. This is observable on the basis of loanwords into Sumerian. For example, the word appearing in Akkadian as *siparru* ‘bronze’ (a word whose ultimate etymology is unknown) appears in Sumerian as *zabar*; the Akkadian reflects an earlier vocalization. Many bisyllabic (and occasionally trisyllabic) roots in Sumerian exhibit the same vowel in each syllable; it is possible that they also are the result of this vowel assimilation. Examples from the core vocabulary of Sumerian include *igi* ‘eye’, *dumu* ‘son’, and *kalam* ‘country, Sumer’. Even certain grammatical suffixes exhibit the same patterning: *-ene*, plural marker for animate nouns; *-eše*, marker of direct speech. It has been suggested that this assimilation of vowels is, in fact, an absolute rule, and that any counterexamples are either loanwords into Sumerian, or the result of compounding (*apin* ‘plow’ is undoubtedly an old loanword; *lugal* ‘king’ is from *lú-gal*, ‘man-great’). However, such a rule of complete vocalic assimilation is intrinsically improbable, and it is difficult to postulate that all exceptions are loanwords. Moreover, some, such as the word typically transliterated as *dumu*, may in fact have been /domu/. There are also grammatical suffixes which do not exhibit this patterning (*-a-ni* ‘his’).

Vocalic assimilation takes place elsewhere, particularly in the prefix chain of the verb. The term “vowel harmony” has sometimes been used, but since this process was nowhere systematic in the language—unlike, say, in Turkish—it is best to use the term assimilation instead. Without going into details, the most prominent cases include:

The vowel of the modal prefixes *ḥa* (affirmative and precative) and *ša* (contrapunctive?) changes in accordance with the following vowel:

<i>ḥa-ba</i>	> <i>ḥa-ba</i>	<i>ša-ba</i>	> <i>ša-ba</i>
<i>ḥa-bí</i>	> <i>ḥé-bí</i>	<i>ša-bí</i>	> <i>ši-bí</i>
<i>ḥa-mu</i>	> <i>ḥa-mu</i> (early texts)	<i>ša-mu</i>	> <i>ša-mu</i>
	> <i>ḥu-mu</i> (later texts)		

The vowel of the modal prefix *nu-* (the all-purpose negative marker) also assimilates before certain conjugation-prefixes:

nu-ba > *la-ba*

nu-bí > *li-bí*

Other instances of regressive assimilation occur, although in less clearly understood contexts. Progressive vocalic assimilation is much rarer. Typically the imperfect aspectual marker assimilates into an /u/ of a verb stem: *-šub-e* > *-šub-u*. Spellings with *-e* (*-šub-e*) also occur; it is not sure if they indicate that this assimilation was optional, or rather represent morphographic spellings which show the underlying full form of the morpheme, even if it was not pronounced as such.

49.12.4. *Vocalic contraction and deletion*

Vocalic contraction is fairly common in Sumerian, but it is difficult to formulate general rules, and in fact it is sometimes difficult to say why contraction was in some places regressive and in other cases progressive. It is frequently assumed that vowels resulting from contraction are long, but again the writing system does not unequivocally show the existence of long vowels.

It is not possible to list here all types of contraction; typical cases include:

/ia/ > /a/: *-a-ni-a* > *-a-na*

/ua/ > /a/: *-ġu₁₀-a* > *ġá*

/ue/ > /u/: *lú-e* > *lú-ú*

The postpositions for the dative, comitative, and ablative cases appear as /ra, da, ta/ respectively after consonants, but as /r, d, t/ after vowels.

49.13. *Future work*

As discussed above, the traditional phonemic inventory cited for Sumerian (including both consonants and vowels) contains only phonemes which are known to exist for Akkadian (except for /ġ/). Although it is clear that the Sumerian phonological system had phonemes which did not exist in Akkadian, there is no unanimity about the inventory of these phonemes, nor about their phonetic nature. Moreover, there has been little investigation of the Sumerian phonological system as a whole. This means that the following questions still need to be resolved: (1) the phonemic inventory and (2) the phonetic nature of the phonemes.

Any investigation of these phonemes can only result from a thorough analysis of all the details of the Sumerian writing system, with all its vagaries. This is a major desideratum in Sumerological studies. Even when this

desideratum is accomplished, however, some Sumerologists believe that we will never be in a position to understand Sumerian phonology, let alone phonetics, to the degree that we understand other ancient languages, such as Akkadian. This is precisely because our knowledge of Sumerian is filtered through Akkadian. Miguel Civil has said, for instance, “the vocalic system of Sumerian will never be satisfactorily recovered.” But other scholars are more optimistic, and believe that further knowledge of the writing system, of Emesal, and of Akkadian scribal practices will enable us to deepen our understanding.

Annotated Bibliography

- Bauer, Josef. 1975. “Zum /dr/-Phonem des Sumerischen.” *Die Welt des Orients* 8:1–9. Reconstructs a consonant not clearly visible in the script.
- Black, Jeremy. 1990. “The Alleged ‘Extra’ Phonemes of Sumerian.” *Revue d’assyriologie* 84: 107–18. Investigates several of the phonemes posited for Sumerian.
- Bobrova, L. V., and A. Yu. Militarëv. 1989. “Towards the Reconstruction of Sumerian Phonology.” In *Lingvisticheskaya rekonstrukciya i drevneyshaya istoriya Vostoka: Materialy k diskussiyam na Mezhdunarodnoy konferencii (Moskva, 29 Maya – 2 Iyunya 1989)*, vol. 1, pp. 96–105. Moscow: Nauka. A comparison of Main Dialect phonology and Emesal phonology, with a view toward reconstructing an earlier stage of Sumerian; also originates the idea that Emesal began as a local dialect which became associated with the goddess Inanna.
- Boisson, Claude. 1989. “Contraintes typologiques sur le système phonologique du sumérien.” *Bulletin de la Société de Linguistique de Paris* 84: 201–33. The typological characteristics of Sumerian.
- Civil, Miguel. 1973. “The Sumerian Writing System: Some Problems.” *Orientalia* 42: 21–34. The problems engendered by the Sumerian writing system for our understanding of Sumerian.
- Falkenstein, Adam. 1959a. *Das Sumerische* (Handbuch der Orientalistik, division 1, vol. 2). Leiden: Brill. Now out of date, but often cited.
- . 1959b. “Untersuchungen zur sumerischen Grammatik. 5. Zum Akzent des Sumerischen.” *Zeitschrift für Assyriologie* 53: 97–105. An early attempt to study stress.
- Hayes, John L. 1990. *A Manual of Sumerian Grammar and Texts*. Malibu: Undena. A pedagogical grammar, with a series of graded lessons teaching the script, vocabulary, and grammar. With bibliography.

- Krecher, Joachim. 1960. "Verschlußlaute und Betonung im Sumerischen." In *Lišan mithurti: Festschrift Wolfram Freiherr von Soden*, pp. 157–97. Kevelaer: Verlag Butzon & Bercker (Alter Orient und Altes Testament 1). Various aspects of Sumerian phonology, particularly stress.
- Lieberman, Stephen. 1977. *The Sumerian Loanwords in Old-Babylonian Akkadian* (Harvard Semitic Studies 22), pp. 24–39. Missoula, Mont.: Scholars Press. Discusses the ways in which the approximate phonetic nature of Sumerian lexemes are recovered.
- . 1979. "The Phoneme /o/ in Sumerian." In *Studies in Honor of Tom B. Jones*, ed. Marvin A. Powell, Jr., and Ronald H. Sack, pp. 21–28. Kevelaer: Verlag Butzon and Bercker (Alter Orient und Altes Testament 203). Interesting for its methodological approach in reconstructing a vowel in Sumerian which was not present in Akkadian.
- Parpola, Simo. 1975. "Transliteration of Sumerian: Problems and Prospects." *Studia Orientalia* 46: 239–58. Despite its title, treats Sumerian phonology in general.
- Schretter, Manfred. 1990. *Emesal-Studien: Sprach- und literaturgeschichtliche Untersuchungen zur sogenannten Frauensprache des Sumerischen*. Innsbruck: Innsbrucker Beiträge zur Kulturwissenschaft Sonderheft 69. The most thorough recent treatment of Emesal.
- . 1993. "Sumerische Phonologie: Zu Konsonantenverbindungen und Silbenstruktur." *Acta Orientalia* 54: 7–30. Discusses the possibility of CCV syllables.
- Sjoberg, Åke, ed. 1984–. *The Pennsylvania Sumerian Dictionary*. Philadelphia: The University of Pennsylvania. There is as yet no up-to-date global Sumerian dictionary; my *Manual* discusses the problems of Sumerian lexicography. Several long-term projects are, however, currently underway. Among the institutions currently engaged in writing a full dictionary of Sumerian is the University of Pennsylvania. The volume treating "B" appeared in 1984, and the first half of "A" in 1993.
- Thomsen, Marie-Louise. 1984. *The Sumerian Language: An Introduction to Its History and Grammatical Structure*. Copenhagen: Akademisk Forlag (Mesopotamia 10). This is the standard reference work. With bibliography.

This page intentionally left blank.

Burushaski Phonology

Gregory D. S. Anderson

University of Chicago

50.1. Introduction

Burushaski, the language of the Burusho, is a language isolate spoken in the Northern Areas region of Pakistan. Attempts have been made to find a genetic connection between Burushaski and many other languages, notably NW Caucasian, Basque, and the Siberian isolate Ket. While some possible lexical correspondences have been noted, the findings are in general of a typological, rather than of a genetic nature.

Though Burushaski is an isolate, there are distinct dialect groups. The major dialect split is between the variants spoken in the Hunza and Nagar regions as opposed to the variant encountered in the Yasin valley. This latter is frequently called Werchikwar in the literature (*Werč* = *Buruš* plus the Khowar suffixes *-ik* ‘place of’ and *-war* ‘language’). The differences among these dialects are discussed in § 50.5.

Burushaski is spoken in the rugged mountain terrain of northern Pakistan and stands at the meeting point of several large language families: Indo-European, Tibeto-Burman, and Altaic, slightly farther to the north. Burushaski has been in contact with the immediately surrounding languages for centuries. To the north live the Wakhi, an Iranian-speaking group whose territory stretches into Tajikistan. Several Wakhi groups live intermixed with the Burusho; in some villages they have even adopted the Burushaski language. To the south, the Dardic Shina predominate. Many Burusho, especially in Nagar, are bilingual in Shina and Burushaski. To the east, Burushaski speakers are in contact to a lesser degree with the Tibeto-Burman Balti. In the Yasin valley, the Burushaski-speakers live intermixed with the Dardic Khowar; the Yasin-Burushaski are all bilingual in Khowar. Finally, in Hunza, especially in the village of Mominabad, the Indo-Aryan speaking

ABBREVIATIONS: B, Burushaski; Dum., Dumaki; Hz, Hunza dialect; Ng, Nagar dialect; Sh, Shina; Skt., Sanskrit; YB, Yasin Burushaski. Morphological classes: hf, human female; hm, human male; x, x-class (roughly equivalent to ‘animate’); y, y-class (roughly equivalent to ‘inanimate’).

Dumaki live in close contact with the Burusho. These latter even call the Doma “Beričo”; virtually all Dumaki speakers are bilingual in Burushaski, and it appears likely that within a few generations they will abandon their traditional language in favor of the prestigious Burushaski language. There are approximately 60,000 speakers of Burushaski, with roughly 1/4 of these speaking the Yasin dialect; the remaining 3/4 are divided nearly evenly among the Hunza and Nagar varieties, with the Nagar dialect enjoying a slight majority of speakers.

While the Burusho are dominant in Hunza, they are less so in Nagar, and they are the minority in Yasin (where Khowar-speakers predominate). Nevertheless, their language remains vital, spoken by all generations in all communities. There is also some evidence that Burushaski-speakers may have previously inhabited a larger area: in Dras, in Baltistan, there is a group of people known as the Brokpa or Bruša; in Ponjal, there are Shina-speaking Burushken.

The earliest reference to Burushaski may be in an old Tibetan source that lists the inhabitants of the Hunza area as the Bru-ža. The first secure attestation of Burushaski is in a wordlist from 1854 by A. Cunningham. Two scholars from the late nineteenth century, J. Biddulph (1880) and G. Leitner (1889), each provided a description of this fascinating language. At the beginning of the twentieth century, the Russian scholar I. Zarubin (1927) and the English administrator D. Lorimer (1935) published fuller descriptions of the Yasin and Hunza/Nagar dialects of Burushaski, respectively. Lorimer’s three-volume grammar still remains the standard. During the present century, several scholars have turned their attention to this intriguing language isolate of northern Pakistan. These include Georg Morgenstierne and Siddheshwar Varma in the 1930s and 1940s, the Soviet scholars Vladimir Toporov and Giorgi Klimov in the 1960s and 1970s, and more recently the German scholar Hermann Berger and Canadians Yves-Charles Morin and Etienne Tiffou.

50.2. The consonantal sounds of Burushaski

The consonantal inventory of B[urushaski] is characterized by a large number of coronal stops and affricates, a phonemic contrast of retroflexion in this area, a phonemic aspiration on voiceless stops and affricates, and a uvular series of stops. In addition, the Hz and Ng dialects of B possess a curious sound whose phonetic realizations vary from a retroflex, spirantized glide,

to a retroflex velarized spirant. This is symbolized [ɣ] in the standard transcription orthography and will receive independent treatment below (see § 50.2.3). From an areal perspective, the phonemic inventory of B is quite similar to that of the neighboring Sh, and shares much in common with most of the surrounding languages.

50.2.1. The consonantal inventory of Burushaski

Table 50-1. Burushaski Consonants

p	t	c	ɬ	č	č̣	k	q
p ^h	t ^h	c ^h	ɬ ^h	č ^h	č̣ ^h	k ^h	q ^h
b	d	z	ɖ	ǰ	ǰ̣	g	ɣ
(f)		s		š	ṣ̌	(x)	h
m	n					ŋ	
w				y	ɣ		
	l	r					

[f] and [x] occur only in loanwords (for example *xalbər* ‘news’, *xanǰər* ‘dagger’, *fəsəl* ‘crops’, *fərs* ‘carpeting’), or as a variant of the aspirated stops [p^h] and [q^h] or [k^h], respectively (see § 50.2.2).

Some examples of phonemic contrasts in B follow:

<i>pəl</i>	‘senseless’	<i>čəp</i>	‘meat’
<i>p^həl</i>	‘grain’	<i>čəp</i>	‘hidden’
<i>bəl</i>	‘narrow’	<i>ǰəp</i>	‘with force, hard’
		<i>ǰəp</i>	‘stay’
<i>təp</i>	‘leaflet’	<i>čəp et-</i>	‘shoot’
<i>dəp</i>	‘strip of teased wool’		
<i>t^həp</i>	‘night’	<i>t^həm</i>	‘king’
<i>ɖəp</i>	‘store-room’	<i>ɬ^həm</i>	‘sweeping’
<i>səŋ</i>	‘light’	<i>cəcəq mənə:s</i>	‘to keep retreating’
<i>šəŋ</i>	‘care, heed’	<i>c^həc^həq mənə:s</i>	‘to trot’
<i>šəŋ</i>	‘boundary’		
		<i>ǰu:k</i>	‘kidney’
<i>d-ə ɣəč-</i>	‘itch[PRES]’	<i>ǰu:k mənə:s</i>	‘touch, reach’
<i>d-ə ɣəč-</i>	‘hide[PRES]’		

<i>ki:l</i>	‘ibex’	<i>kʰʌn</i>	‘castle’
<i>kʰi:l mʌna:s</i>	‘to see’	<i>gʌn</i>	‘road’
		<i>ɣʌn</i>	‘gold tax’
<i>qʌm</i>	‘pit’		
<i>kʌm</i>	‘little’	<i>ʃum</i>	‘precipitous’
		<i>ʃun</i>	‘intelligent’
<i>dʌli</i>	‘upper’		
<i>dʌri</i>	‘window’	<i>ʃʌm</i>	‘completely’
		<i>ʃʌŋ</i>	‘war’
<i>d-ʌɣʌɣʌs</i>	‘to itch’		
<i>d-ʌɣʌɣʌs</i>	‘to hide’		

50.2.2. *Phonotactics and allophony*

Consonant clusters at the beginning of words in B are extremely limited. Most of the words with initial clusters are loanwords, although some may not be, e.g. *pra:q* ‘completely’, *trʌqo* ‘grasshopper’, and *praiyiš* ‘clean’. The clusters that may be found word-initially in B are limited to unaspirated anterior stop + [r], that is *br-*, *pr-*, *dr-*, and *tr-*; *tʰrʌq* ‘cleft’ and *pʰra:* ‘stone in ring’ are the only examples of an initial aspirated stop + [r] cluster attested in B. In addition, a very limited number of initial Cy- clusters are attested, most of which are onomatopoeia, e.g. *čyu* ‘chirping’, *q(i)yu* ‘shout’; in at least one instance, however, a Cy- cluster is not onomatopoetic: *pyu(wʌn)* ‘a little’. Note that initial clusters are not attested in any B verb-root.

Medially, numerous two-member clusters are found, consisting of a continuant (usually [l], [n], or [r], though sometimes [s] or [ʃ] followed by an obstruent). Thus, no stop and/or affricate clusters are tolerated. Geminates occur mainly in loanwords, and are infrequent.

Final clusters—basically identical to those encountered word-medially—are likewise frequently encountered, e.g. *gunc* ‘day’, *kʰʌsk* ‘girth’, *mʌnč* ‘adze’, *hərʃ* ‘plough’, *ti:ʃk* ‘dagger’. In the Hz dialect these have sometimes been simplified, for example Hz *ɣʌs* ‘sister’ vs. YB *ɣʌst*.

In addition, the sounds *-z*, *-g*, *-b*, *-d*, *-ɖ*, *-ʃ*, *-ʒ*, *-ɣ*, *-x*, *-f*, *-w*, and *-y* are not found word-finally in native B vocabulary.

Also, not only are initial clusters not found in verb roots, the following sounds are likewise never encountered root-finally in B verbs: {-*p*, -*č*, -*ć*, -*m*, -*w*, -*ŋ*}. In fact, most B verbs end in *-n*, *-l*, *-r*, *-ɣ*, *-y*, *-s*, some end in *-t* or *-q*, and very few in *-t*, *-k*, *-c*. Thus, no B verb root is labial-final.

Another phonotactic restriction seen in B prevents homorganic sequences of glide + vowel. Thus, **yi-* and **wu-* are not tolerated in B words.

Few sounds exhibit regular allophony, most simply exhibiting freely varying pronunciation in certain environments. The voiced fricative [ɣ] is generally heard as either an affricate [gɣ-] or the voiced uvular stop [ɢ] in onset position; in some speakers it sounds in general more like uvular [ɤ] than velar [ɣ]. In addition, within the phonological system of B, [ɣ] functions as the voiced counterpart of [q], see § 50.4.1. Also, in many speakers the velar sounds alternate freely with the corresponding uvular consonants before [a], for example *qa:* for /ka:/, *gula:p* < Urdu *gula:b* ‘rose’.

Other, analogous “strengthening” of consonants is also encountered, though with less consistency than the above example. [x] is often pronounced as an affricate, and frequently varies with [q], [qʰ] or [kʰ] in syllable-onset position, as in *xat* ~ *kʰat* ‘down’, *xam* ~ *qʰam* ~ *kʰam* ‘cooked vegetables’, or *xərc* ~ *qərc* ‘clapping hands’. This strengthening is also seen in certain morphological environments as well, for example in perfect participle formation (see § 50.4.1). Similarly, [f] is often pronounced as an affricate [pf], and varies with [pʰ] syllable-initially.

Aspirated consonants exhibit the reverse process of “weakening” in syllable-initial position. Thus the aspirated consonants *qʰ* and *pʰ* alternate with [x ~ qx] and [f ~ pf], respectively, in onset position, e.g. *qʰap* ~ *xap* ‘tinder’, *kʰu:rk* ~ *xu:rk* ‘broken straw’, *pʰat* ~ *pfat* ~ *fat* ‘release’, *kafan* ~ *kapʰan* ‘shroud’. Medially, aspirates are of infrequent occurrence in the Hz dialect, and may be related to stress placement (the aspirates do not occur in post-stress position); this is not true of Ng B, however (see § 50.5.1).

Another allophonic tendency is the alternation of [x] and [h] (word-initially) before [u]: *xumu:in* ~ *humu:in* ‘sonless man’, *xurgas* ~ *hurgas* ‘thick’. The correspondence of standard B *q/qʰ* : YB *h*- seen in such words as Hz *quronč*, Ng *qʰaronč*, Yasin *huronč* ‘cloud’ may be related to this.

[h] is not uncommon in initial position, but is frequently lost in intervocalic position (following a stressed vowel), see § 50.3.3. In addition, [h] occurs word-finally mainly in loanwords. However, words that end in a short stressed vowel are sometimes pronounced with an *h*-like voicelessness, e.g. *saʰ* ‘sun’, *buaʰ* ‘cow’. Note that voiceless final vowels are also encountered in some Sh dialects (e.g. Pa:las or Kohistyō:).

The sounds [j] and [jʰ] alternate freely with the corresponding fricatives [ž] and [žʰ], respectively. The voiced counterpart of [c] is [z], as *[j] is lacking in B, a trait it shares with surrounding languages such as Sh and Kashmiri. An example of this is seen in *hu:nc* ‘arrow’ > *hu:nze* ‘arrows’.

In some words, the voiceless coronal fricatives alternate with the corresponding homorganic affricate; for example *sum* ~ *cum* ‘female (domestic

animal)’ or $c^h\lambda n\check{\varsigma} \sim c^h\lambda n\check{\varsigma}$ ‘every time, whenever’. In some instances, this may be the result of inter-dialect borrowing (see § 50.5.1).

The B sound [l] tends to be devoiced when it stands before a voiceless consonant. B [r] is rare in initial position, especially so in non-loanwords. In addition, although [l] and [r] are phonemically distinct in B, some words exhibit variation between the two sounds: *bil* ~ *bir* ‘full’ or *-lpur* ~ *-rpur* ‘eyelash’ (the latter may be due to assimilation or dissimilation, however). Note that there are sporadic [l] : [r] correspondences in the various dialects of B, see § 50.5.

B [n] shows a tendency towards assimilation of secondary articulatory features, such as retroflexion. Thus [n] ~ [ɳ] / __ Ç; similarly before [ʃ]/[ç], [n] alternates in pronunciation with [ɳ]. In addition, in word-final position, nasal sounds vary with one another, for example Hz. *un* ~ *uŋ* ‘you (sg.)’, *-ri:ŋ* ~ *-ri:n* ‘hand’. However, [ŋ] is not merely **ng* in B, as is evident from the minimal pair *-ŋgi* ‘in front of’ vs. *-ŋi* ‘beard’ (note that **ngg*—or any three-member cluster—is an impossible sequence in B).

50.2.3. The problem with [ɣ]

The B sound [ɣ] is a distinctive feature of the Hz and Ng dialects. This sound is lacking in YB (Werchikwar); words in Hz that end in [-ɣ] often correspond to final glottal closure (ʔ or *h*), long vowel, or sometimes [ɣ] in YB, e.g. YB *baʔ/bah* vs. Hz *bay* ‘millet’, YB *bu:-* (PRES stem *buč-!*) vs. Hz. *buy-* ‘(y.sg.) dry up’ (see § 50.5.2). It occurs only in post-vocalic position, i.e., non-initially. Its pronunciation varies widely; in word-final position it is realized as a retroflex glide [ɣ]. Following a velar or uvular sound before [u], this sound is realized as a strongly labialized retroflex continuant. In addition, word-medially [ɣ] is frequently heard as a retroflex [ɣ̣] or [ɣ̤]. Historically in loanwords, B [ɣ] comes from Sanskrit [-ḍ]/[-ṭ], as in B *payo* ‘wedge’ < Skt. *paṭaka* or *baḷum* ‘mare’ < Skt. *vaḍabaḥ* (> **vaḍam* > Sh *ba:m*). The language of the Ḍumaki (Beričo clan) also has this [ɣ] sound. Frequently, Ḍum. [ɣ] corresponds to B [ɣ]; however, in both Ḍum. and B, [ɣ] alternates with [ɣ̣] in certain words, for example Ḍum. *nila:yo* ~ *nila:yo* ‘forehead’ or B *m̄ayo* ~ *m̄ayo* ‘fine’, and there is sometimes also a correspondence Ḍum. [ɣ̣] : B [ɣ], as in Ḍum. *baḷum* ‘mare’ vs. B *baḷum* or Ḍum. *š̄amey* vs. B *š̄amey* ‘yoke-pin’ (as well as Ḍum. [ɣ̣] < Skt. [ḍ], as in *ta:ya* < Skt. *taḍa* ~ *ta:la* ‘palm of hand’). As was stated above, sometimes Hz/Ng [ɣ] corresponds to YB [ɣ̣] as well, e.g. Hz *d̄ay* ‘stout, fat’ vs. YB *d̄aỵanum* (h/x.pl.) ‘fat’. Note also the B variants *-mu:ruỵas* ~ *-mu:ruṭ̄(in)as* ‘to pollard’.

Within B itself, an instance of [ɣ] patterning like a retroflex [ɣ] is seen in the morphophonological changes associated with the B present stem formation (see § 50.4.2). A further role of [ɣ] within the phonological system of B is a phenomenon that might be called “cerebral assimilation” that was characteristic of the speech of Lorimer’s informants; today, this is either highly restricted or lacking altogether. This caused the retroflex feature associated with [ɣ] to regressively assimilate to a preceding [g], [q], or [ɣ]. An example of this phenomenon is *giɣas* ~ *giyas* ‘to enter’. The issue is, however, more complicated than this example might lead one to believe. The [+“retracted”] feature seems not to spread to a voiceless [k] seen in the perfect participle form of the above verb *niki:m* ‘entered’. This lack of “cerebral assimilation” may possibly also be true of aspirates; however, relevant examples are lacking in my materials, and therefore it is unknown whether these sounds are affected or not. A further example of “cerebral assimilation” is seen in present stem formation, discussed in § 50.4.2; also, data from YB suggest that this process may have been characteristic of pre-B (see § 50.5.3).

50.3. The vowels of Burushaski

B has a basic five-vowel system, with distinctive vowel length.

Table 50-2. Burushaski Vowels

i		u	i:		u:
	e	o		e:	o:
	a			a:	

Examples of phonemic vowel contrasts in B: (short vowels) : (long vowels) *-lanc* ‘leg’ vs. *-lta:nc* ‘eyebrow’; *bat* ‘flat stone’ vs. *ba:t* ‘cooked rice’; *hiʃ* ‘breath’ vs. *hi:ʃ* ‘abundance’; *dir* ‘boundary’ vs. *di:r* ‘overhanging rock’; *bun* ‘high country’ vs. *bu:n* ‘boulder’; *ɣun* ‘block’ vs. *ɣu:n* ‘quail’; *mən* ‘who’ vs. *mɛ:n* ‘old’; *e:garimi* ‘he let him play’ vs. *egarimi* ‘he played’; *doras* ‘to grind’ vs. *do:ras* ‘to fall’. Minimal pairs involving [ɛ:] ([e:]) or [o:] are of infrequent occurrence. Long vowels only appear in stressed syllables in B.

In careful speech, most sequences of two vowels are syllabic. In actual usage, however, a variety of diphthongs are encountered. As a general tendency, [i] becomes non-syllabic [y] before a stressed syllable, e.g. *ayotimi* ‘he did not make them’; in other positions, individual speakers vary as to

whether diphthongs are formed, for example *dírias* ~ *díryas* ‘ripen’. Sometimes, morphological environments require a specific realization, for example, the morphologically-conditioned vowel coalescence phenomena discussed in § 50.3.3. However, diphthongs are formed in one phonological environment in Burushaski. Thus, [a] becomes [ai] or [au] before an initial *y* or *w*, respectively, in “sandhi” position: *hai yAkAl* < *ha y-* or *xau wašimi* ‘threw down’ < *xa w-*.

50.3.1. Suprasegmentals

In addition, there are non-segmental accentual phenomena in B. Varma (1941: 133) describes these intonational phenomena as representing a rising and falling tone; modern investigators, however, e.g. Tiffou (personal communication), Berger, and Morin (personal communication), consider this to be a difference of moraic stress: that is, long vowels may receive stress on either the first mora or the second, corresponding to Varma’s falling and rising tones, respectively. These intonational phenomena are phonemic in B, and may distinguish both lexical and grammatical forms. For example, *íi* ‘his son’ vs. *íi* ‘himself’ or *óomaltaras* ‘to envelop them’ vs. *oómaltaras* ‘to not envelop’. In addition, the former type is sometimes heard with a lowered pitch on the first mora, for example *c^hùum* ‘difficult’ or *raĵâaki* ‘forced labor’; note that diminutives are generally associated with this intonational pattern, e.g. *šon* ‘blind’ vs. *šòon* ‘somewhat blind’ or *ġak* ‘attached’ vs. *ġâak* ‘somewhat attached’. YB exhibits the same intonational phenomena as the standard Hz/Ng varieties. However, the moraic stress difference is less pronounced, especially in diminutive formations.

Although long vowels are found only in stressed syllables in underived lexical items, in most derived forms the placement of stress in B is morphologically, not phonologically, determined. Thus, in bound noun roots (except those beginning in [-ʌ] or -IC-) stress falls on the first syllable of the root; in the case of the exceptions, stress falls on the prefix: *a-tátas* ‘my palm’, *a-súsun* ‘my elbow’, *a-q^hát* ‘my mouth’, but *á:skus* ‘my mother-in-law’ (< -ʌskus), *á-lpur* ‘my eyebrow’, *á-ltumul* ‘my ear’. An example of a morpheme attracting stress comes from the causative *á(s)-*, for example *ʌbilas* < *bɛlas* ‘put on clothes’ (on the change of *ɛ* > *i*, see § 50.3.3). As was stated above, long vowels are realized only in stressed position, e.g. *ȝaí:ȝ* ‘grape’, *gušpú:r* ‘prince’, or *ȝé:niš* ‘gold’. A bisyllabic sequence of a heavy syllable (heavy = CV: or CVC) followed by a short one generally is pronounced with initial stress, for example *hú:nze* ‘arrows’ and *ȝá:mo* ‘ice’. In addition, lexical

items consisting of a short syllable followed by a long one are described as oxytones, but frequently receive stress on the final heavy syllable: *šapík* ‘bread’ or *halýur* ‘horse’. However, a certain portion of these lexical items are pronounced with initial stress: *kʰápun* ‘spoon’, *hámal* ‘neighbor’. Note that stress placement is distinctive in B, e.g. *báre* ‘of the valley’ vs. *baré* ‘behold!’ or *duđúr* ‘apricot species’ vs. *duđur* ‘small hole’.

In derived forms intervocalic [h] is frequently lost except when in onset position of a stressed syllable. Thus, the causative of *huru:las* ‘sit’ is *áurutas* ‘make sit’ (on the shortening of the stem vowel, see § 50.3.3), and the negative imperative of *her-* is *ówar* ‘don’t weep’, but the perfect participle form is *nuhéř*.

50.3.2. Allophony

Although distinctive in the language, the length of a vowel in a given word may vary from speaker to speaker, for example *gupaltij* ~ *gupa:ltij* ‘trousers’. In addition, the phonemic status of long vowels is less strong in YB than in the Hz and Ng varieties.

However, certain vowels do exhibit fairly regular allophony. For example, short /a/ is pronounced as [ə] before [r], as in *gar* ‘marriage’ (cf. *ga:r* ‘giddiness’) or *šikar* ‘hunting’, as [æ] before or after palatals (y, š, č, čʰ, j), as in *qænč* ‘cave’ or *šæni* ‘vegetable garden’, while in most other instances it is realized as [ʌ].

In addition, unstressed /i/ and /u/ tend to lax to [ɪ] and [ʊ], respectively. Before a uvular consonant or [r], [u] is frequently lowered to [o], as in *moqʰakin* ‘her daughter-in-law’ (< *mu-*). In addition, in Ng, unstressed [u] “umlauts” to [ü] before a palatal consonant or [i], for example *amüli* ‘where’ or *šüčo* ‘festival’. Vocalic phenomena in YB are discussed in § 50.5.2.

[ɛ] is the major allophone of /e/. In many words they are in free variation, e.g. *bepay* ~ *bepay* ‘yak’; *de:mi(y)-* ~ *dɛ:mi(y)-* ‘spill, pour out’; in other words, the form has one vowel or the other in its basic pronunciation, for example *deskəɾay-* ‘heat’ or *di(y)eray-* ‘become weak’. This distribution is frequently related to stress placement. Though not absolute, word-medial [ɛ] receives stress, whereas final [-e] may be unstressed.

[o] is found only in stressed position.

Individual speakers vary as to whether diphthongs are formed in quick speech. Various combinations in stressed position, however, give rise to a range of coalescences. For example, the sequence *óu* > *óo*, while the sequence *uó* > *oó*, as in *óoruṭimi* ‘he set them down’ (< **o-u*) vs. *doór-* (< **du-or*) ‘fall

down'. The sequence *éa* > *ée*, while *áe* > *áa*, e.g. *jaméen* 'a cross-bow' (< *é-a*) and *háa* (< *há-e*) 'of the house'. In unstressed position, the underlying sequences /ai/ and /au/ are frequently realized as [ei] and [ou], respectively, as in *akóomei baa* 'you can't' or *ouláat* 'descendants'.

50.3.3. *Vocalic phenomena*

As was stated above, the placement of stress is often morphologically conditioned. In addition, certain other vocalic phenomena are conditioned by morphological factors. Two such phenomena are the coalescence of [a] and [u] to [ɔ] and [o:] (*oó*). The former is created in certain morphological environments when [u] stands before [a], for example in the affixation of the second person singular possessive prefix *gu-*, as in *gɔkin* (< *gu-akin*) 'your liver' or *gɔi* (< *gu-ai*) 'your daughter' vs. *guʃ* 'your neck'. Long [o:] is formed when the prefix *a-* is added to stressed *u*-initial stems, for example *o:mus* (< *a-umus*) 'my tongue' or *o:ʃis* (< *a-uʃis*) 'my foot' vs. *anɪ* 'my beard'. An example of this latter change in conjunction with *h*-loss comes from *o:lʃa* < *a-hulʃa* 'don't mount'. Another coalescence encountered is [a] and [i] forming [e:] (sometimes [ɛ:]), as in *e:rɕaba* 'I die' vs. *guirɕuma* 'you die'.

When the negative prefix *a-* is added to stems that begin with vowels, either the vowels coalesce as described above, or an epenthetic [y] is inserted to break up the vowel cluster. The epenthesis occurs when [a] is attached to non-high vowels, while if it is prefixed to stems that begin with high vowels, coalescence occurs. For example, we find *aiya:rin* 'don't send me' < *a + a:rin*, *a(i)yeti* 'don't do it' < *a + eti*, *ayo:tuma kɛ* 'if they did not make them' < *a + o:tuma*, but *e:u:ɕa* 'I do not give him' < *a + i + u:ɕa*.

Another morphologically conditioned vowel phenomenon might be dubbed "causative shortening/tensing." When causative verbs are formed, a long stem vowel frequently shortens, while [ɛ] tenses and raises to [i], e.g. *wa:las* 'to lose' > *ɔspalas* (on the change of [w] > [p], see § 50.4.1), *ba:ltas* 'wash' > *ɔbaltas*, or *cu:yas* 'take away' > *ɔcuyas*; *-deɭas* > *ɔdilas* 'make jump', *bɛlas* > *ɔbilas* 'make put on'.

There is also a sporadic syncopation of vowels associated with certain morphological formations. This deletes a (usually high) vowel in the penultimate syllable of certain stems when deriving various morphological forms. These include the perfect participle *nV-* and denominal verb-formation. Examples of this syncopation include *gu:ʃuginas* 'to consult' > *nuku:ʃkin* (on the change of [g] > [k], see § 50.4.1) and *ʃiqər* 'shame' > *-ʃqərɔs*. Note the YB form *nuhrut* 'having sat' < *huru:t* (cf. Hz *nu:ruʃ*).

50.4. Consonant alterations

Certain alternations of consonants operate in morphologically, rather than phonologically, conditioned environments. In most cases it is likely that this results from the morphologization of previously allophonic distinctions in pre- or proto-B, with the phonological conditioning lost due to subsequent change. These morphophonological rules include one of the language's most characteristic alternations: the devoicing (or general strengthening) of certain consonants in intervocalic position (§ 50.4.1). Another morphologically conditioned process is palatalization (§ 50.4.2) associated with present-stem formation and certain plurals. In § 50.4.3, a type of regressive "vowel harmony" or "vowel-copy" associated with certain morphological formations is discussed. The various manipulations of aspiration seen in the morphology are addressed in § 50.4.4.

50.4.1. Intervocalic devoicing

B exhibits a curious devoicing (strengthening) of certain consonants in intervocalic position in certain morphological environments. When various prefixes are added, the stem-initial consonant devoices. These prefixes include the formants of the negative stem in *a-*, the perfect participle in *nV-* (see § 50.4.3), and the preverb *dV-*. The consonants that undergo this alternation are [g], [b], [d], [ɣ], and [ʃ], which become [k], [p], [t], [q], and [č] respectively, e.g. *bɛ:l-* 'put on clothes' > *ʌpɛ:l* 'NEG'; *nupɛ:l*; *ɣɪlti:r-ʌs* > *nɪqɪltir* 'be extravagant, show off'. This devoicing is especially typical of the second person prefix *gu-*, e.g. from *d-mana* 'be born', we get *dukumana* 'you were born'; from *dil(in)* 'hit' we get *nukudilin* 'having hit you'. Note the form *atica ba* 'I did not bring (him)' < *di:ca* shows that the devoicing also affects the *dV-* preverb, which itself triggers this same rule.

The causative affix *ʌs-* also triggers this devoicing (and strengthening) rule. However, with this affix, sometimes a progressive assimilation of the devoicing is encountered; thus, from *dʌɣai-ʌs* we get *ʌstʌqai-ʌs* 'conceal' and *babʌl* gives *d-ʌspʌpʌl-ʌs* 'carry away'.

In addition, in some of these morphological environments, [w] and [h] are similarly strengthened to [p] and [k], respectively. Examples of this include *wərc* 'right-side out' > *apa:rc* 'reverse, wrong-side out' (cf. *ʌspərlʌs* 'to make tired' < *-wərlʌs*) or *henas* 'to know' > *akenas* 'not to know' (cf. *ʌskir* 'father-in-law' < *hir* 'man'); but, note *nuhən* 'having known'. This strengthening of continuants was noted as a phonetic variation in § 50.2. Many of those same stems exhibit strengthening in these morphological environments. For

example, from *-xəras*, *ixəras* ‘cut up’ we get the perfect participle *niqər*; from *xəratas* ‘stick to, adhere’ we find the causative *ΛqərΛtas*.

50.4.2. Palatalization

Palatalization of certain consonants is characteristic of the formation of the present stem and certain plurals. Verb forms fall into two classes, those formed from the past base, and those formed on the basis of the present stem. Frequently, the latter is derived from the former. In many instances, this is simply via affixation. In stems ending in certain consonants, however, the old affix has coalesced with the stem-final consonant, causing palatalization; thus, these verb stems form their present stem via the morphologically conditioned palatalization rule. This palatalization rule triggers the following alternations: [k]/[s] > [š], [c]/[t] > [č]/[š], [y] > [č], [ɣ] > [č], and [n] > [y]/[iy]. Thus, we get

PAST	PRES	
<i>yan-</i>	<i>yay-</i>	‘take’
<i>yec-</i>	<i>yesh-</i>	‘see’
<i>-sərk-</i>	<i>-sərš-</i>	‘allow’
<i>girat-</i>	<i>giraš-</i>	‘dance’
<i>et-</i>	<i>eč-</i>	‘do’
<i>d-ayas-</i>	<i>d-ayaš-</i>	‘laugh’
<i>man-</i>	<i>ma(i)y-</i>	‘become’
<i>d-ΛγΛy-</i>	<i>d-ΛγΛč-</i>	‘hide’
<i>d-ΛγΛy-</i>	<i>d-ΛγΛč-</i>	‘itch’

In many instances, it is precisely this palatalization that signals the morphological category and meaning of a form, e.g. *du:šΛm* ‘I (will) bring them’ vs. *du:cam* ‘I brought them’.

The [n] > [y] alternation probably went through a stage of palatal **ñ*; note the correspondence Hz [y]- : YB [ny]- in the word *ya* ~ *nya* ‘bear’ (see § 50.5.2). The final two examples above (‘hide’ and ‘itch’) represent a minimal pair. From these examples, it appears that the retroflex articulation of [ɣ] is some kind of autosegmental feature that is preserved even when the segment it is associated with becomes something else. Further evidence for the independence of this feature comes from the curious retracted velar sounds described by Lorimer and mentioned in § 50.2. For a discussion of this, see § 50.5.3.

The second morphological feature that palatalization is associated with in B is plurality. When the plural affix *-o* is attached to certain stem-final con-

sonants, these undergo palatalization. The changes encountered are [n] > [ɲ], [s] > [ʃ], and [c] > [č]:

SG.	PL.	
<i>daman</i>	<i>damayo</i>	‘master’
<i>ɣin</i>	<i>ɣiyo</i>	‘thief’
<i>hiles</i>	<i>hilešo</i>	‘boy’
<i>balas</i>	<i>balašo</i>	‘bird’
<i>huye:lterc</i>	<i>huye:lterčo</i>	‘shepherd’
<i>gu:ic</i>	<i>gu:ičo</i>	‘Wakhi’

Most investigators assume this palatalization reflects a plural suffix *-yo in pre-B, and therefore to have a natural motivation for this phonetically unnatural rule. The form *ya* ‘cow’, plural *ɣaiyo*, is usually considered to be evidence for this formant.

50.4.3. Regressive vowel harmony

In addition to the two important morphologically conditioned alternations adduced above, there are also other morphophonological phenomena characteristic of B. One of these is a type of regressive vowel harmony or vowel-copy. The vowel in the common preverb *dV-* and the perfect participle formant *nV-* generally appears as a copy of the vowel of the following syllable, when attached to consonant-initial forms. With the former, this is almost always the case, while with the latter, the vowel-copy is sometimes not encountered. From the verb *d-cas* ‘bring to s.o.’ and *n-dil(in)* ‘hit’ we get

<i>dacas</i>	‘bring to me’	<i>nadilin</i>	‘having hit me’
<i>duku:cas</i>	‘bring to you’	<i>nukudilin</i>	‘having hit you’
<i>di(:)cas</i>	‘bring to him’	<i>ni:dil(in)</i>	‘having hit him/(it)’ (–anim)
<i>dumu:cas</i>	‘bring to her’	<i>numu:dilin</i>	‘having hit her’
<i>dimi:cas</i>	‘bring to us’	<i>nimi:dilin</i>	‘having hit us’
<i>dama:cas</i>	‘bring to you pl.’	<i>nima:dilin</i>	‘having hit you pl.’
<i>du:cas</i>	‘bring to them’	<i>nu:dilin</i>	‘having hit them’

Note that the perfect participle formant does not have vowel-copy in the second plural form. Note also that both *dV-* and *nV-* trigger the devoicing rule described in § 50.4.1.

50.4.4. Aspiration

There is lexicalized evidence that stress (and aspiration) may have performed at least a limited derivational function in the history of the language. There are transitive/intransitive pairs that appear to be derivationally related to

each other by a differing stress pattern and the presence or absence of aspiration on a consonant: *d-c^higinlas* ‘hang’ (intr.) vs. *d-cikinlas* ‘hang’ (tr.); *di-nc^hi:ras* ‘be spread (out)’ vs. *di-ncirlas* ‘spread out’ (tr.). In the former cases (the intransitive forms), there is aspiration on the consonant and stress follows it; in the transitive forms, there is no aspiration, and stress precedes.

In addition, a number of morphological processes result in deaspirating an aspirated stem-consonant. Thus, with the *nV*- participle, an aspirated consonant deaspirates: *-č^hilas* ‘give’ > *ničin*, *guc^hər* ‘walk’ > *nukucər*, *-p^h/flətər* ‘flay’ > *niplətər*. In Ng, however, deaspiration is not operative: *nič^hi* ‘having given’ or *nukuc^hər* ‘having walked’. Similarly, causatives cause deaspiration in Hz but not the Ng dialect: Hz *a:kucəras* ‘make one go’ vs. Ng *a:guc^həras*. This deaspiration in Hz is probably related to a shift in the stress of the word.

50.5. Dialects

There are distinct dialect forms of the B language. In the preceding pages, the term B has been mainly (though not exclusively) limited to the standard Hz dialect. In the following sections we enumerate some of the more salient phonological features of the dialects of B. In § 50.5.1, the phonological characteristics distinguishing the Hz and Ng varieties are discussed, while section § 50.5.2 addresses the more divergent language of the YB (Werchikwar). Finally, § 50.5.3 addresses some evidence from the lexicon in these dialects that might represent lexicalized remnants of earlier phonological phenomena.

50.5.1. Nagar and Hunza

Though Lorimer discusses regional variation in B, the best comparative study of Hz and Ng B is Varma (1941). Hz and Ng B exhibit dialectal variation mainly in the lexicon, though phonological and morphosyntactic differences are also encountered. From a phonological standpoint, Ng B is quite similar to the Hz dialect. Though individual speakers exhibit variation in pronunciation within a given dialect, there are nevertheless some general correspondences between the two dialects. For example, Ng frequently has a diphthong where the corresponding Hz form shows a simplex vowel:

Nagar	Hunza	
<i>olteik</i>	<i>oltik</i>	‘both of them’
<i>miušindaro</i>	<i>mišindaro</i>	‘our wives’

<i>bium</i>	<i>bim</i>	‘they were’
<i>teilər</i>	<i>tɛlɐr</i>	‘thither’
<i>ɭuyər</i>	<i>oyər</i>	‘my husband’
<i>a:us</i>	<i>o:s</i>	‘my wife’
<i>ša:uq</i>	<i>šo:k</i>	‘noose’
<i>ɭu-</i>	<i>o:-</i>	‘NEG.prefix’

Note that the characteristic devoicing associated with the negative prefix *a-* in Hz is not triggered by the Ng formant *ɭu* or its Hz cognate *o:-*.

The Ng dialect sometimes has nasalized vowels corresponding to oral vowels in Hz; this is also true of nasalized vowels in words borrowed from Sh.

<i>čĩlto</i>	<i>čilto</i>	‘third day after tomorrow’
<i>māš</i>	<i>maš</i>	‘massage’

Another frequently encountered vowel correspondence is Ng [-u] : Hz [-o] in final position:

<i>hɪpultu</i>	<i>hɪpulto</i>	‘day after tomorrow’
<i>hunzu</i>	<i>hunzo</i>	‘Hunza’
<i>mutu</i>	<i>muto</i>	‘now’

Whether this is the result of some kind of assimilation process in Ng is unknown at present. Another vowel correspondence between Ng and Hz B is encountered in unstressed position; here we find Ng [ɔ] corresponding to Hz [-u]. This may be due to the adjacent uvular sound (cf. similar phenomena in YB).

<i>dɔyoi</i>	<i>duyui</i>	‘noon’
<i>yɔski</i>	<i>yuski</i>	‘freshly kneaded dough’

Finally, there is a sporadic correspondence of Ng [æ] : Hz [a], as in

<i>gæn</i>	<i>gan</i>	‘road’
<i>æpi</i>	<i>api</i>	‘is not’

Consonantal correspondences between Ng and Hz B are likewise not exceptionless. Nasals in final position show a wide range of variation, as they do among individual speakers within a given dialect. Medially, Hz frequently has [-ŋ-] corresponding to Ng ø:

<i>um</i>	<i>un/uŋ</i>	‘you (sg.)’
<i>dum</i>	<i>dun</i>	‘a little’
<i>p^hin</i>	<i>p^him</i>	‘foam’
<i>hi:ɬi</i>	<i>hiŋai</i>	‘doors’
<i>ruɬɐnc</i>	<i>ruŋɬɐnc</i>	‘pasture-lands’
<i>tɪayo</i>	<i>tiŋayo</i>	‘eggs’

The [m] in Ng could of course be due to labial assimilation to the preceding [u]. Also, a sporadic correspondence of Ng [m] : Hz [b] has been noted. This could be an assimilation to the nasal in the same or adjacent syllable—a similar phenomenon has operated in the course of the development of numerous Turkic languages.

Nagar	Hunza	
<i>man etas</i>	<i>ban etas</i>	‘kiss’
<i>nimaŋte</i>	<i>nibaŋte</i>	‘on reaching’

In final position, Ng [-c] frequently corresponds to Hz [-s]; in addition, word-medially, the Hz cluster [-st-] corresponds to Ng [-sc-].

<i>ɣərkaɫc</i>	<i>ɣərkaɫs</i>	‘lizard’
<i>balac</i>	<i>balas</i>	‘bird’
<i>descəl-ɫas</i>	<i>destəl-ɫs</i>	‘to awaken’
<i>ɣisca</i>	<i>ɣista</i>	‘leavened bread’

In addition, one sporadically encounters a correspondence of Ng [-ʃ] : Hz [-s], most often in coda position.

<i>aʃkil</i>	<i>askil</i>	‘face’
<i>mɔʃ</i>	<i>mɔs</i>	‘flood’
<i>ʃal</i>	<i>sar</i>	‘pit’

Word-medially, Ng often shows an aspirated consonant where Hz has an unaspirated one:

<i>Alt^hər</i>	<i>Altər</i>	‘twenty’
<i>baɫ^hin</i>	<i>baɫin</i>	‘leg’
<i>ɣaɫt^həl</i>	<i>ɣaɫtəl</i>	‘deer’
<i>dik^hətɫas</i>	<i>dikətɫas</i>	‘to be entangled’

In loanwords (especially those originating in English) Hz has a word-initial aspirate corresponding to an unaspirated consonant in Ng. Conversely, in some words of unknown origin, Ng exhibits a word-initial aspirate, whereas the corresponding Hz form has no aspiration. Note also the lack of deaspiration associated with various morphological formations in the Ng dialect (see § 50.4.4).

<i>pulis</i>	<i>p^hulis</i>	‘police’
<i>koɫ</i>	<i>k^hoɫ</i>	‘coat’
<i>q^hɫeŋɟ</i>	<i>qɫeŋɟ</i>	‘sword’
<i>t^hariŋ</i>	<i>tariŋ</i>	‘skin-bag for milk’

Numerous sporadic correspondences within the velar/uvular places of articulation have been noted between the Hz and Ng dialects of B. Among these are Ng [ɣ-] Hz [k-] or Ng [k^h-] : Hz [g-] word-initially, Ng [-ɣ-] : Hz [-q-] medially, and Ng [-x] : Hz [-q] in final position.

<i>ɣʌruɣo</i>	<i>kaɽuɣo</i>	‘moa bird’
<i>ɣuʈul</i>	<i>kuʈul</i>	‘mud-covered wood basket’
<i>k^hutɛ</i>	<i>gutɛ</i>	‘this (near)’
<i>dicciɣinʌs</i>	<i>dicciqinʌs</i>	‘to hang’
<i>eʃoɣun</i>	<i>eʃuqun</i>	‘younger’
<i>rʌx</i>	<i>rʌq</i>	‘wish’
<i>lʌx</i>	<i>lʌq</i>	‘naked’

For the most part there is an exact correspondence of liquid sounds between these dialects; however, sporadically, a correspondence of Ng [l] : Hz [r] is attested. In addition, Ng occasionally loses [l] in a medial cluster before an obstruent, while this cluster is preserved in Hz.

<i>ʃʌl</i>	<i>sʌr</i>	‘pit’
<i>bilʌgo</i>	<i>birʌgo</i>	‘colt’
<i>ɖʌŋ ləʈʌs</i>	<i>ɖʌŋ rəʈʌs</i>	‘female goblin’
<i>ɛ:tirʌs</i>	<i>ɛ:ltirʌs</i>	‘to show’
<i>deɟʌba</i>	<i>deɟʌba</i>	‘I strike’

Note, however, Ng *tur biʃʌɣəs* Hz *tul biʃʌɣəs* ‘break into a house’.

50.5.2. Werchikwar

The best and most complete study of the language of the YB (Werchikwar) is Berger (1974). YB exhibits a greater degree of phonological divergence from the Hz/Ng standard than either of the latter two do from each other. As a general tendency, YB shows closer affinity to the phonological system of the Ng dialect.

In terms of vowel correspondences, YB vocalism is most often identical to that of Hz/Ng in cognate forms; vocalic length is only marginally phonemic in YB. In a number of forms, we find YB [ɛ] : Hz/Ng [i].

Yasin	Hz/Ng	
<i>čɛn</i>	<i>čin</i>	‘bird’
<i>sɛs</i>	<i>sis</i>	‘persons’
<i>hɛk</i>	<i>hik</i>	‘one’/‘full’
<i>bɛs</i>	<i>bis</i>	‘fat’
<i>ɛmɛʃ</i>	<i>ɛmis</i>	‘finger’
<i>dʌsɛn</i>	<i>dʌsin</i>	‘girl’

Conversely, we also find a number of correspondences of YB [i] : standard B [ɛ] before [l].

Yasin	Hz/Ng	
<i>sil</i>	<i>sɛl</i> (Ng - <i>ɪ</i> -)	‘needle’
<i>dil</i>	<i>dɛl</i>	‘oil’

In at least a few cases, there is a correspondence of YB [ʌ] : Hz/Ng [i]. Occasionally, YB itself shows vacillation between [ʌ] and [i] among individual speakers. More frequent however, is the correspondence of YB [ɛ] : B [ʌ]. Note that [i] : [i] and [ʌ] : [ʌ] are most frequent of all.

<i>hʌɛs</i>	<i>hileɛs</i>	‘boy’
<i>cʰʌl/cʰil</i>	<i>cʰil</i>	‘water’
<i>yuyɛŋ</i>	<i>guyʌŋ</i>	‘hair’
<i>hareʃ</i>	<i>harʌʃ</i>	‘urine’
<i>yuryɛʃ</i>	<i>gurʌʃ</i>	‘feces’

Another correspondence encountered between the YB and B proper is [u:] : [i:] before [r].

<i>ʌsu:r</i>	<i>ʌsi:r</i>	‘near’
<i>hu:ri</i>	<i>hi:ri</i>	‘men’

In a few cases, YB has the diphthong [yɛ-/yɪ-] corresponding to Hz/Ng [i:-].

<i>-yɛ/yɪ(:)</i>	<i>-i:</i>	‘son’
<i>-yɛk</i>	<i>-i:k</i>	‘name’
<i>-yu:r</i>	<i>-i:r</i>	‘die’

Similar to standard B, in the environment of palatal consonants, or preceding a stressed [i], YB [a] is frequently realized as [e].

<i>ʃeni/ʃani</i>	‘garden’	<i>jenjer/ʃanjer</i>	‘shackles’
<i>-hekin/-hakin</i>	‘learn’	cf. <i>hakin-</i>	‘teach’

Within YB itself, a stressed *u* is frequently realized as *o* before a uvular, and sporadically in a few other words. Conversely, unstressed *o* frequently raises to *u* in YB, except when it is adjacent to a velar-uvular continuant or immediately before a stressed *o*. In addition, from the forms just cited we see that unstressed *a* in YB shows a tendency to “umlaut” to *e* before stressed *i* (less often *e*). Finally YB *i* alternates with *e* either after a uvular or before *s*, *ʃ*, *ʒ*, *r*, or *l*, for example *yearum* ~ *yiarum* ‘utensils’ or *pʰareʃ* ~ *pʰariʃ* ‘duck’. In YB, round vowels tend to front after palatal consonants, e.g. *žü* ‘apricot’, *süva* ‘good’, *čüst* ‘beautiful’; *šön* ~ *šin* ‘blind’.

The consonantism of YB is in general quite similar to that of standard B. However, the Yasin dialect lacks the characteristic [y] sound found in Hz and Ng. Frequently, YB cognates exhibit a long vowel where the Hz/Ng form has the sequence [Vy]. However, in a few cases, YB has a final [-h] corresponding to Hz/Ng [-y].

<i>do:mo</i>	<i>doypa</i>	‘right’
<i>ya:mo</i>	<i>gaypa</i>	‘left’
<i>baḥ</i>	<i>baɣ</i>	‘millet’
<i>bu:-</i> (PRES <i>bu:č!</i>)	<i>bu(:)ɣ-</i>	‘dry up’ (y.sg. subj.)

Secondly, Yasin possesses a voiceless, spirantized [ɬ], which in some speakers varies with [θ] in pronunciation; though lacking in Hz and Ng dialects, the adjacent Dardic language Khowar also possesses this sound. In older sources (e.g. Leitner 1889) YB final [-ɬ] corresponds to Hz/Ng [-ll]. (Because we cannot be sure that in Leitner’s pre-IPA account orthographic <ll> does not simply represent [ɬ] as in Welsh, we cannot definitively state that Hz/Ng did not possess a now-lost [ɬ].)

<i>taɬ</i>	<i>tall</i>	‘birch’
<i>paɬ</i>	<i>yall</i>	‘shadow’

In some morphological environments YB [ɬ] becomes [l], e.g. *tu:maɬ* > *tumaliŋ* ‘ear’. Note that this is morpholexically conditioned, not phonologically, cf. *bɛɬis* ‘sheep’ or *cuɬum* (Hz/Ng *cu:m*) ‘heavy’. In medial position YB often has a homorganic nasal plus stop sequence where the corresponding Hz/Ng form has only a nasal.

<i>yendiš</i>	<i>gyɛniš/ɣeniš</i>	‘gold’
<i>geŋgi</i>	<i>gaŋgi</i>	‘axe’

Other correspondences of nasals between the dialects are YB [ny] : Hz/Ng [y], as in *nya* : *ya* ‘bear’ or YB *nyaɬ* : Hz/Ng *yall* ‘shadow’; on occasion, the correspondence YB [m] : Hz/Ng [b] is attested in medial position, e.g. *taɬmaŋ* : *taɬbaŋ* ‘bridle’. Again, this may be due to an anticipatory assimilation of nasality.

As is evident from the forms just given, a frequently encountered correspondence is YB [ɣ-] : standard B [gɣ-/g-] in initial position.

In word-final position, some clusters have undergone simplification in Hz/Ng but are preserved in YB, for example Yasin *yaɬst* : standard B *yaɬs* ‘sister’.

Within YB itself initial [y-] alternates with [ž-], as in *ya-/ža-* in *yabap* ~ *žabap* ‘my father’. In addition, the voiced palatal affricates frequently are pronounced as the corresponding homorganic fricative in word-final position; thus, [j] and [ɟ] > [ž] and [ʒ], respectively. Similarly, the voiceless aspirated [pʰ] alternates freely with [f]; in final position, [pʰ] always weakens to [f]. The voiceless coronal affricates assimilate to their voiced counterparts when following [n]; thus [č] , [č̣], and [c] > [j], [j̣], and [z], respectively. YB [r] is realized as [š] before [č], as in *bašč* ‘bridge’ < **barč*.

50.5.3. Cerebral assimilation

Various phenomena suggest that a process of “cerebral assimilation” has been operative during the course of the development of B. This “retracted” feature spreads from [ɣ] to a preceding voiced velar or uvular sound (see § 50.2.3). In addition, verb stems ending in -ɣ behave like stems ending in -y plus an autosegmental “[+retracted]” feature in present stem formation (see § 50.4.2). Also, further evidence for a “retroflexing” autosegment in B comes from YB forms such as *çiş* ‘mountain’ or *çumuş -man-* ‘catch a cold’ (< **çiş*, **çumuş*); note *çot* ‘heap’ and *daɬ* ‘hole’ from **çot* and **daɬ*, respectively. Note also the correspondence Yasin *šenj*: Hz/Ng *sinç* ‘beam’. The Yasin dialect has lost the characteristic [ɣ] sound, though “cerebral assimilation” was apparently still active during the development of this dialect, cf. the root *bu*: ‘dry up’ PRES stem *bu:ç-*. Finally, there is evidence that this “[+retracted]” feature may have played a limited derivational function in the history of Burushaski. There are some verb stems that are subcategorized for a singular y-class (inanimate) actant, either subject or object depending on the specific verb’s valence. Some of these verbs have [ɣ] in the form subcategorized for a y-singular actant while the corresponding y-plural form has [y], for example *bu:ɣas* ‘(y.sg.) dry up’ (PRES *bu:ç-*) vs. *bu:yayas* (PRES *buyač-* ‘(y.pl.) dry up’ or *du:ɣas* ‘(y.sg.) melt’ vs. *du:yayas* ‘(y.pl.) melt’.

Finally, as mentioned above, a peculiar aspect of the lexicon is the presence of verbs subcategorized for actants of a given class (one of the four possible noun classes). As in other languages, sometimes these stems stand in a suppletive relationship to one another. In others, they are clearly derivationally related synchronically. However, for the present purposes, still other sets of these stems seem to be derivationally related to each other diachronically via a stem-initial consonant alternation. The most frequently encountered of these correspondences are h/x-class y-/w- vs. y-class g-/b-. Thus, for example, we find:

h/x-class	y-class	
<i>ɣanas</i>	<i>ɣanas</i>	‘take’
<i>ɣatamuras</i>	<i>ɣatamuras</i>	‘knead, wring’
<i>yo:las</i>	<i>bɛlas</i>	‘put on’
<i>ya:ltas</i>	<i>ba:ltas</i>	‘wash’
<i>wašiɣas</i>	<i>bišayas</i>	‘throw’
<i>-wa:las</i>	<i>balu:yas</i>	‘become lost’
<i>-walas</i>	<i>bala:s</i>	‘fall’

References

- Berger, Hermann 1974. *Das Yasin-Burushaski*. Wiesbaden: Harrassowitz.
- . forthcoming. *Grammatik der Burushaski Sprache* [tentative title].
- Biddulph, John. 1880. *Tribes of the Hindoo Koosh*. Calcutta: GPO.
- Cunningham, A. 1854. *Ladakh: Physical, Statistical, and Historical*. London.
- Klimov, G. A. & D. I. Edel'man 1970. *Iazyk burushaski*. Moscow: Izdatel'stvo "Nauka".
- Leitner, Gottlieb. 1889. *Hunza and Nagyr Handbook*. Calcutta: GPO.
- Lorimer, D. L. R. 1935. *The Burushaski Language*, 3 vols. Oslo: H. Aschehoug & Co. (W. Nygaard).
- Morgenstierne, Georg. 1945. "Notes on Burushaski Phonology." *Norsk Tidsskrift for Sprogvidenskap* 13: 61–95.
- Toporov, V. N. 1970. "About the Phonological Typology of Burushaski." In *Studies in General and Oriental Linguistics Presented to Shirô Hattori*, ed. Roman Jakobson and Shigeo Kawamoto, pp. 632–47. Tokyo: TEC.
- Varma, Siddheshwar. 1941. "Studies in Burushaski Dialectology." *Journal of the Royal Asiatic Society of Bengal. Letters* 7: 133–73.
- Zarubin, I. 1927. "Veršikskoe narečie kandžutского iazyka." *Zapiski Kollegi Vostokovedov* 2/2: 275–364.